No. 86119-6

IN THE SUPREME COURT OF THE STATE OF WASHINGTON

STATE OF WASHINGTON,

Respondent,

v.

BRYAN EDWARD ALLEN,

Petitioner.

ON REVIEW FROM THE COURT OF APPEALS, DIVISION ONE

BRIEF OF *AMICI CURIAE*
COLLEGE AND UNIVERSITY PROFESSORS
JENNIFER DEVENPORT, JENNIFER E. DYSART,
GEOFFREY LOFTUS, STEVEN D. PENROD,
NANCY K. STEBLAY and GARY L. WELLS,
IN SUPPORT OF DEFENDANT BRYAN EDWARD ALLEN

/8557   ALLEN, HANSEN & MAYBROWN, P.S.
One Union Square
600 University Street, Suite 3020
Seattle, Washington 98101
Telephone: (206) 447-9681

MILBANK, TWEED, HADLEY & McCLOY LLP
1 Chase Manhattan Plaza
New York, New York 10005
Telephone: 212-530-5088
Of Counsel to Attorney for *Amici Curiae*
College and University Professors

TABLE OF CONTENTS

TABLE OF AUTHORITIES

Page(s)

INTRODUCTION

This Court has an opportunity to survey the large body of

persuasive scientific research on the unreliability of eyewitness

identifications developed since the decision by this Court in *State v.*

Laureano, 101 Wn.2d 745, 767-68, 682 P.2d 889 (1984), *overruled on*

other grounds by State v. Brown, 111 Wn.2d 124, 132-33, 761 P.2d 588

(1988). The jury instructions at issue in *Laureano*, were modeled on non-

scientific instructions in a 1972 case, *United States v. Telfaire,* 469 F.2d

552 (D.C. Cir. 1972) Questions involving eyewitness identification, which

were largely guesswork in the 1970's have been elucidated by rigorous,

peer-reviewed research that is now widely accepted and has been hailed as

the "gold standard" of scientific research. As the New Jersey Supreme

Court observed when it recently had the same opportunity to review the

science of witness identification: "Experimental methods and findings

have been tested and retested, subjected to scientific scrutiny through

peer-reviewed journals, evaluated through the lens of meta-analyses, and

replicated at times in real-world settings." *State v. Henderson,* 208 N.J.

208, 27 A.3d 872, 916 (2011). Indeed, this Court has already recognized

some of the key scientific concepts discussed in this brief, such as the

fallibility and malleability of memory, and the importance of estimator

variables such as "weapon focus." *State v. Cheatam*, 150 Wn.2d 626, 648, 81 P.3d 830 (2003).

This Court has also recognized that juries should be advised about "special" witnesses, such as accomplices, that require a "special kind of attention" from juries. *State v. Carothers*, 84 Wn.2d 256, 268, 525 P.2d 731 (1974) (*en banc*). As this Court recognized in *Carothers*, courts have expertise in evaluating special witnesses that the ordinary citizen cannot be expected to have. *Id.* After decades of hearing scientific experts testify on eyewitness identification issues, courts understand them better than jurors commonly do.

While the question on appeal involves only a specific subset of the mental processes of forming, storing and retrieving memories, and only one of the mechanisms used to safeguard against erroneous eyewitness identification, we present research on the full range of topics at issue in the case below to permit a comprehensive review by this Court of this large and complex body of science.

STATEMENT OF THE CASE

The key relevant trial testimony is follows:

The victim/witness was white and the perpetrator was African-American. (21st 8:12-16, 9:6-7.) The perpetrator was wearing a hooded sweatshirt, a baseball cap and sunglasses. (21st 9:21.) The perpetrator threatened to

kill the witness and showed him what appeared to be a gun in his waist-band. (21st 30:3-5.) The witness was scared (21st 11:1-8.) and "disturbed" and "terrified." (20th 6.) He described the perpetrator as around his own height and weight (21st 32:20-33:17), which is five feet, nine inches tall and 210 to 220 pounds. (21st 32:20-33:12.) An African-American male in the area who was significantly taller and heavier than described, wearing a hooded shirt, a baseball cap and sunglasses, was detained. (20th 8:2-6.) While the witness was with a police officer, a radio call came in with the information that a suspect was "at gunpoint." (20th 6:15.) The police officer told him they were driving to see a suspect and to see if he could identify him. (20th 11:15-18; 10/21/09 23:17-25.) The officer told a fellow officer to "uncuff" the detainee, who was several blocks away. (20th 7:5-7; 21st 50:14-16.) The record does not reflect the witness was out of earshot for this conversation. The witness saw the detainee surrounded by multiple police officers. (20th 11:5-7.) The police asked the detainee to put on his sunglasses on to assist with the identification (21st 16:15-17; 70:16-17) and to pull his cap down on his head, "as he had been at the crime." (21st 70:15-17.) After the identification, the police "immediately" handcuffed the detainee. (21st 71:7-9.) The witness testified at trial: "if I were to walk down the street right now I probably wouldn't recognize him." (21st 34:12-13.)

ARGUMENT

Over the past three decades, there has been an explosion of research in the eyewitness identification field, which now contains the largest and most rigorous body of scientific research of all law-related social science fields. Social scientists, including Amici, have conducted thousands of eyewitness identification experiments. These experiments demonstrate how certain factors, both in isolation and in tandem with other factors, can erode the reliability of eyewitness identification evidence. This research has been published in hundreds of articles in a range of peer-reviewed psychological journals. By relying on the scientific method and using sound experimental designs, eyewitness identification researchers have been able to ensure the internal validity[1] of their research findings.[2]

Publication of scientific findings in peer reviewed journals which disclose the experimental designs and analytical methods is an important consideration in determining the validity of the findings. The peer review process is a method of quality control that ensures the validity and reliability of experimental research. *Henderson*, 27 A.3d at 892-93 (citing Mal-

---

[1] Internal validity refers to the reliability of a laboratory study's research method, and hence the reliability of its result. A study is internally valid if it has a control group, avoids confounding factors, randomly assigns participants, eliminates the possibility of experimenter bias to the greatest extent possible and does not arbitrarily omit data. *See* Malpass et al., *The Need for Expert Psychological Testimony on Eyewitness Identification*, in *Expert Testimony On The Psychology Of Eyewitness Identification* 3, 12-13

pass, supra note 2). By employing the scientific method to eyewitness identification research, generations of researchers have ensured that the research comported with the methodological requirements of the most demanding peer review journals in the field of psychology. This Court, for example in *State v. Cheatam*, has acknowledged the force and weight of this research.[3]

The scientific rigor of eyewitness identification research is established not only by its quantity and its quality, but also by the consistency of its findings on particular variables, which are best captured by meta-analytic reviews within the academic literature. Meta-analyses combine data sets from large numbers of published studies performed by different researchers in different labs under different circumstances (and can also include the results of field studies), and convert them into a common metric known as the 'effect size.'[4]

By merging several data sets that address the same question, *id.* at 893-94, meta-analyses can ascertain the mean effect size across a number of studies, serving as a more powerful estimate than the effect size demon-

---

(Brian L. Cutler ed., 2009); Levine & Parkinson, *Experimental Methods in Psychology* (1994). Additionally, any scientific inquiry must be falsifiable to be valid.
[2] Malpass *et al.*, *supra* note 2 at 12-13.
[3] "[T]there are numerous studies showing that contrary to many jurors' beliefs upon questioning, it is more difficult for people of one race to identify people of a different race." *Cheatam*, 150 W.2d at 645-46, 81 P.2d at 840.
[4] Shapiro & Penrod, *Meta Analysis of Facial Identification Studies*, 100 Psychol. Bull. 140, 140 (1986).

strated by a single study under a single set of assumptions and conditions. By this means, a meta-analysis can detect consistent and significant patterns of witness behavior across studies. The more consistent the conclusions from aggregate data regarding the magnitude of impact of an independent variable, the greater confidence scientists have in those conclusions.[5] *Id.* at 893. By "systematically us[ing] prior research findings to generate a fairly precise estimate of the effect sizes detected in a body of research, [meta-analyses] . . . provide a succinct summary of the status of scientific research in a particular domain."[6]

I.   THE FALLIBILITY OF MEMORY

Cognitive psychologists have long "established that when we experience an important event, we do not simply record it in our memory as a videotape recorder would."[7] Rather, what is perceived and stored in memory is "often incomplete or distorted as a result of the individual's state of mind or the nature of the event observed."[8]

Scientists analyzing the nature of memory have focused on its three discrete stages: (1) the acquisition or encoding stage, when a witness perceives an event and information is thereby entered into the mem-

---

[5] *See also id.* (detailing the findings of several meta-analyses).
[6] *See* Penrod & Bornstein, *Generalizing Eyewitness Reliability Research* 2 *The Handbook of Eyewitness Psychology: Memory for People* 529, 535 (L. Lindsay et al. eds., 2007) (detailing the findings of several meta-analyses).
[7] Loftus et al., *Eyewitness Testimony: Civil and Criminal* § 2-2, at 12 (4th ed. 2007).

ory system; (2) the retention or storage stage, the period between acquisition and the witness's attempt to recall the information; and (3) the retrieval stage, when the witness attempts to recall the stored information.[9] "This three-stage analysis is central to the concept of human memory," and "[p]sychologists who conduct research in this area try to identify and study the important factors that play a role in each of the three stages."[10] Those psychologists, including Amici, have identified in particular numerous factors that may adversely affect an eyewitness's memory at each stage. At the acquisition stage, memory is subject to both event-specific variables (such as the presence of a weapon at the crime scene) and witness-specific variables (such as race).[11] At the retention stage, additional factors such as the passage of time or post-event information may contaminate the witness's memory.[12] At the retrieval stage, a witness's memory can be impacted by the environment in which the retrieval is taking place and the wording of questions being asked.

Building on this body of research regarding the nature of memory generally, scientists have conducted many empirical studies—most using

---

[8] Brigham et al., *Disputed Eyewitness Identification Evidence*, 36 Ct. Rev. 12, 13 (1999).
[9] *See* Loftus et al., *supra* note 7, at 13.
[10] *Id.*
[11] *Id.*
[12] *Id.* Because memory is subject to many sources of contamination, researchers have recommended that it be regarded as similar to trace evidence from a crime scene, such as a fingerprint or hair sample. *See, e.g.*, Wells, *Eyewitness Identification: Systemic Reforms*, 2006 Wis. L. Rev. 615, 622-23 (2006).

controlled experimental methods—that document the adverse impact of various factors on the accuracy of eyewitness identification.

II.   EYEWITNESS RESEARCH

Researchers agree that the accuracy of eyewitness identifications is significantly affected by two types of factors: "estimator variables," which relate to the eyewitness's direct experience of the event, and "system variables," which relate to investigative procedures following the event.

A.   Estimator Variables Present in this Case

Estimator variables are factors beyond the control of the criminal justice system.[13] They can include factors related to the incident, the witness, or the perpetrator and can affect an eyewitness' ability to perceive and remember an event.

1.   Own-race or Cross-race Bias

Extensive empirical research demonstrates that people are better able to remember faces of their own race than those of a different race, a phenomenon known as "own-race bias" or "cross-race bias." Own-race bias is a very robust effect, found across a variety of racial groups,[14] age

---

[13] *See* Wells, *Applied Eyewitness-Testimony Research: System Variables and Estimator Variables*, 36 J. of Personality and Social Psychol. 1546, 1546 (1978).

[14] *See e.g.* Chance & Goldstein, *The Other-Race Effect and Eyewitness Identification, in Psychological Issues in Eyewitness Identification* 153-176 (Ludwig Siegfried et al. eds., 1996); Meissner & Brigham, *Thirty Years of Investigating the Own-Race Bias in Memory for Faces: A Metaanalytic Review*, 7 Psychol. Pub. Pol'y & L. 3-35 (2001).

groups[15] and recognition tasks.[16] Therefore, as the New Jersey Supreme Court found, "cross-racial recognition continues to be a factor that can affect the reliability of an identification." *Henderson*, 27 A.3d at 907.

There are at least three meta-analyses involving this phenomenon, and all support the conclusion that other-race recognition is less reliable than same-race recognition. Most recently, a 2001 meta-analysis of research involving laboratory-based face-recognition[17] found that the odds of a <u>mistaken</u> identification are 1.56 times higher with other-race faces than with same-race faces. Similarly, the odds of a <u>correct</u> identification are 1.4 times higher for own-race face as compared with performance on other-race faces, and the odds of a correct decision are 2.2 times higher in same-race faces, as compared to other-race faces.

Two earlier meta-analyses looked specifically at experiments involving white and African-American subjects.[18] The unequivocal conclusion of both of these earlier meta-analyses was that both white and African-American subjects evidenced an overall tendency to recognize faces of their own race better than other-race faces. Though scientists do

---

[15] *See e.g.* Corenblum & Meissner, *Recognition of Faces of Ingroup and Outgroup Children and Adults*, 93 J. Experimental Child Psychol. 187–206 (2006).

[16] *See, e.g.,* Lindsay *et al., Other-Race Face Perception,* 76 J. Applied Psychol. 587-89 (1991); Meissner & Brigham, *supra* note 20, at 3-35.

[17] Meissner & Brigham, *supra* note 14, at 15, 21 (spanning 39 research articles and nearly 5,000 participant witnesses).

not fully understand the origins of the differential encoding systems that cause the own-race effect, research suggests that the effect does not seem to be related to prejudice.[19]

2. Disguise

Disguises may include hats, glasses, wigs and facial hair, as well as masks. *Henderson*, 27 A.3d at 907. Research has shown that the encoding process for storing information about a face is impaired when a perpetrator is simply wearing a hat[20] or sunglasses.[21]

One experiment measured the effects of such "disguise" on subsequent identification accuracy by using a perpetrator in a staged setting who wore a knit pullover cap covering his hair and hairline in some cases and not in others. Identification accuracy was appreciably reduced for witnesses in the disguise condition, from 45 percent accuracy in the no-hat condition, to 27 percent in the disguise condition.[22]

---

[18] Bothwell et al., *Cross-Racial Identification*, 15 Personality & Soc. Psychol. Bull. 19 (1989); Anthony *et al.*, *Cross-Racial Facial Identification: A Social Cognitive Integration*, 18 Personality & Soc. Psychol. Bull. 296 (1992).

[19] *See* Meissner & Brigham, *Thirty Years of Investigating the Own-Race Bias in Memory for Faces: A Meta-Analytic Review*, 7 Psychol. Pub. Pol'y & L. 3, 7 (2001) (citing numerous studies).

[20] Cutler, *et al.*, *Improving the Reliability of Eyewitness Identification: Putting Context Into Context*, 72 J. Applied Psychol. 629 (1987).

[21] *See* Hockley, *et al.*, *Shades of the Mirror Effect: Recognition of Faces With and Without Sunglasses*, 27 Mem. Cogn. 128, 130 (1999) ("Faces, particularly if they appear wearing sunglasses, may also engender strong, accurate, and confident feelings of familiarity in the absence of recollection.").

[22] Patterson & Baddeley, *When Face Recognition Fails*, 3 J. Experimental Psychol. Hum. Learning & Memory 406, 410, 414 (1977).

3.   Weapon Focus

When a visible weapon is used during a crime, it can distract a witness and draw attention away from the culprit. *Henderson*, 27 A.3d at 904. The phenomenon of "weapon focus effect" occurs when the presence of a weapon interferes with an eyewitness's ability to encode a perpetrator's face.[23] The weapon-focus effect interrupts the victim's attention to the perpetrator's face and may also inhibit the memory trace by affecting long-term memory formation.[24]

A meta-analysis of nineteen weapon-focus studies that involved more than 2,000 identifications found an average decrease in accuracy of about 10 percent when a weapon was present.[25] In a separate study, half of the witnesses observed a person holding a syringe in a way that was personally threatening to the witness; the other half saw the same person holding a pen.[26] 64 percent of witnesses from the first group misidentified a filler from a target-absent lineup, compared to 33 percent from the second group.[27]

---

[23] Loftus et al., *Some Facts About "Weapon Focus,"* 11 L. & Hum. Behav. 55 (1987).
[24] Steblay, *A Meta-Analytic Review of the Weapon Focus Effect*, 16 L. & Hum. Behav. 413 (1992).
[25] *Id.*
[26] Maass & Kohnken, *Eyewitness Identification: Simulating the "Weapon Effect"*, 13 L. & Hum. Behav. 397, 401-02 (1989).
[27] *See id.* at 405.

Weapon focus can also affect a witness's ability to describe a perpetrator. A meta-analysis by amicus Dr. Nancy K. Steblay of ten studies showed that "weapon-absent condition[s] generated significantly more accurate descriptions of the perpetrator than did the weapon-present condition."[28] A recent dissertation study at John Jay College (conducted by an active duty Connecticut police chief) used a videotaped robbery and found that in a no-weapon condition, witnesses were able to correctly identify the target 78 percent of the time.[29] When a weapon was implied (by the perpetrator waving his hands around in his pocket), accuracy dropped to 55 percent and when a weapon was actually shown, accuracy dropped to 33 percent.

The duration of the crime is also an important consideration. Weapon-focus studies speak to real-world "situations in which a witness observes a threatening object . . . in an event of short duration." *Henderson*, 27 A.3d at 905. Thus, when the interaction is brief, the presence of a visible weapon can affect the reliability of an identification and the accuracy of a witness' description of the perpetrator. *Id.*

---

[28] Steblay, *supra* note 24, at 417.

[29] DeCarlo, *A Study Comparing the Eyewitness Accuracy of Police Officers and Citizens* (2010) (unpublished Ph.D. thesis, City University of New York).

4. Stress

A witness to a violent crime experiences heightened stress, a defensive reaction with clear physiological responses (*e.g.,* acceleration in heart rate, increased blood pressure and muscle tone).[30] By comparison, a person experiencing moderate stress would undergo none of these physiological changes. Research has demonstrated that a witness's high level of stress reduces the accuracy of identifying a perpetrator.[31]

In 2004, Amicus Dr. Steven Penrod and other researchers examined all of the studies on the relation of high stress to accuracy in identifications.[32] The meta-analysis concluded that high stress reduced correct identification rates by one-third, from 59 percent to 39 percent, compared to identification rates involving low stress,[33] supporting the hypothesis that high levels of stress negatively impact accurate recall and correct identification rates,[34] even while moderate levels of stress can actually improve cognitive processing.

To illustrate, a study at the "Horror Labyrinth," a London Dungeon tourist attraction, used self-report measures of anxiety, validated against

---

[30] Deffenbacher et al., *A Meta-Analytic Review of the Effects of High Stress on Eyewitness Memory*, 28 L. & Hum. Behav. 687 (2004) (citing
Psychol. Rev. 185 (1984)).
[31] Tredoux *et al., Eyewitness Identification,* in *Encyclopedia of Applied Psychology* 875, 878 (Charles Spielberger ed., 2004).
[32] Deffenbacher *et al., A Meta-Analytic Review of the Effects of High Stress on Eyewitness Memory*, 28 L. & Hum. Behav. 687 (2004).
[33] *Id.*

heart-rate changes.[35] Tourists encountered a "scary person" while slowly walking around the labyrinth later, they were tested to see if they could identify the scary person from a 9 person photo-array. Only 18 percent of those with higher anxiety were able to identify the scary person/culprit, compared with 75 percent of the witnesses experiencing lower anxiety.

B. System Variables Present in this Case

Scientists have developed a large body of research on the extent to which variables outside the estimator can affect witnesses' storage and retrieval of memories, leading them to "misremember" both faces and the circumstances in which they saw them.

1. Showups

A "showup" procedure is essentially a one-person line-up: a single suspect is presented to a witness to make an identification. Often showups involve persons who were apprehended because they were near the scene of the crime and matched the overall description given by the eyewitness.[36] By their nature, showups implicitly indicate a belief that a suspected perpetrator has been identified. *Perry v. New Hampshire*, 565 U. S. ____ (2012) (slip op. at 13).

---

[34] *Id.*

[35] Valentine & Mesout, *Eyewitness Identification Under Stress in the London Dungeon*, 23 Applied Cognitive Psychol. 151 (2008).

[36] *See* Dysart et al., *Show-ups: The Critical Issue of Clothing Bias*, 20 Applied Cognitive Psychol. 1009 (2006).

The relative unreliability of a showup procedure is evident when compared with an unbiased lineup procedure, which is truly a test of the match between the suspect's appearance and the witness's memory of the perpetrator. In a properly carried-out lineup, an innocent suspect is falsely identified only if, by chance, he matches the witness's memory of the perpetrator better than do the five fillers. In a showup procedure, by contrast, a positive identification of the suspect should depend on a match between the suspect's appearance and the witness's memory of the perpetrator, but it can very well depend as well on irrelevant factors, including the witness's expectation that a detained person is the perpetrator, any pressure on the witness to make a positive identification, a natural proclivity to say "yes" or "no" in such a situation, and a victim's desire for an arrest. Showups cannot be conducted blind or double-blind. In fact, as discussed below, showups increase the risk that witnesses will base identifications more on similar distinctive clothing than on similar facial features.[37] As a result, there is no principled manner to assess how much credence to put in a positive identification in a showup procedure.

A meta-analysis of 12 studies involving 3,013 participants found that in target-present presentations, the showup and lineup will produce

---

[37] *See id.* 1019; *see also* Yarmey et al., *Accuracy of Eyewitness Identifications in Show-ups and Lineups*, 20 L. & Hum. Behav. 459, 461, 470 (1996); Gonzalez et al., *Response Biases in Lineups and Showups*, 64 J. Personality & Soc. Psychol. 525 (1993).

approximately the same results (46 percent vs. 45 percent correct identifi-

cations), but that dangerous false identifications of innocent suspects are

much higher in showups than lineups (23 percent vs. 10 percent).[38]

Unlike properly designed and conducted lineups, showups fail to

provide a safeguard against witnesses with poor memories or those in-

clined to guess. *Henderson*, 27 A.3d at 903. A proper lineup tests the

match between a witness's memory of a perpetrator and the suspect's ap-

pearance: a mistaken identification could be the selection of either the

suspect or a filler. During a showup, however, every mistaken identifica-

tion will point to an innocent suspect. *Id.* Researchers have found that

"false identifications are more numerous for showups [compared to line-

ups] when an innocent suspect resembles the perpetrator."[39]

2.   Clothing Bias

Clothing bias occurs when a witness makes an identification from

a procedure (such as a showup) where the clothing that the perpetrator was

wearing and the clothing worn by a suspect in the procedure are the same

or very similar.[40] Clothing bias has been found to decrease identification

accuracy. A study in 1996 found that when a similar-looking (innocent)

---

[38] Steblay et al., *Eyewitness Accuracy Rates in Police Showups and Lineup Presenta-
tions: A Meta-Analytic Comparison*, 27 L. & Hum. Behav. 523 (2003).
[39] *See id.* at 523 (conducting meta-analysis).

16

suspect was present in a showup wearing the same clothing that was worn by the target, false identifications were higher than when the innocent suspects wore different clothing.[41] In short, the risk to an innocent suspect who has a similar appearance to a perpetrator may be increased by clothing bias.[42]

Research has shown that approximately 50 percent of all information reported by eyewitnesses relates to a perpetrator's clothing[43] and are often more distinctive than physical descriptions of the person.[44] The participants in a study conducted by Amica Jennifer Dysart and others, made significantly more false identifications of innocent suspects wearing the same distinct clothing as that worn by a target with similar appearance.[45] When a witness provides a general physical description combined with a detailed clothing description, law enforcement is likely to select a suspect based on a clothing match if such a person is located near the scene of a

---

[40] Lindsay et al., *Do the Clothes Make the Man?: An Exploration of the Effect of Lineup Attire on Eyewitness Identification Accuracy,* 19 Canadian J. Behav. Sci. Special Issue: Forensic Psychol. 741 (1987).

[41] Yarmey et al., *Accuracy of Eyewitness Identifications in Show-ups and Lineups,* 20 L. & Hum. Behav. 459 (1996).

[42] Dysart et al., *Show-ups: The Critical Issue of Clothing Bias,* 20 Applied Cognitive Psychol. 1009 (2006).

[43] Lindsay et al., *Do The Clothes Make The Man? An Exploration Of The Effect Of Lineup Attire On Eyewitness Identification Accuracy,* 19 Can. J. Behavioral Sci. 464-78 (1987); Lindsay, et al., *Default Values in Eyewitness Descriptions: A Problem for the Match-to-Description Lineup Foil Selection Strategy,* 18 L. & Hum. Behav., 527-41 (1994).

[44] Dysart et al., *Mugshot Exposure Prior to Lineup Identification Interference, Transference, and Commitment Effects,* 86 J. Applied Pyschol. 1280 (2001).

crime. When a suspect in such a circumstance is placed in a showup, clothing bias presents an even greater risk of false identification.[46] As a result, if a person who resembles the perpetrator is apprehended near the scene of the crime, is placed in a showup, and is wearing distinct clothing similar to that described by the eyewitness, the likelihood of false identification is considerable.[47]

3.   Pre-Identification Instruction Bias

There is a broad consensus for the practice of preceding all identification procedures with instructions to the witness that the perpetrator may or may not be in the lineup or array and that the witness should not feel compelled to make an identification.[48] Scientists agree that without an appropriate warning, witnesses may misidentify innocent suspects because they look more like the perpetrator than do the other lineup members.[49]

A 1981, study found that the failure to warn witnesses that the perpetrator may or may not be present resulted in 78 percent of witnesses

---

[45] Dysart et al., *Show-ups: The Critical Issue of Clothing Bias*, 20 Applied Cognitive Psychol. 1009 (2006).

[46] *Id.*

[47] *Id.*

[48] *Henderson*, 27 A.3d at 897 (calling the practice "uncontroversial").

[49] There is a significant decrease in mistaken identification when witnesses are warned that the perpetrator may not be present in the lineup. *See, e.g.*, Steblay, *Social Influence in Eyewitness Recall: A Meta-Analytic Review of Lineup Instruction Effects*, 21 Law & Human Behav. 283 (1997); Clark, *A Re-examination of the Effects of Biased Lineup Instructions in Eyewitness Identification*, 29 Law & Human Behav. 395 (2005).

making an identification – even in a target-absent lineup – whereas by giving the warning reduced the number of mistaken identifications to 33 percent.[50] Since then two separate meta-analyses have found that there is a significant decrease in mistaken identification when witnesses are warned that the perpetrator may not be present in the lineup.[51] Thus, as *Henderson* found, "The failure to give proper pre-lineup instructions can increase the risk of misidentification." *Henderson*, 27 A.3d at 897. Pre-identification instruction bias is plainly a risk in the case of a showup where the witness is told that he will be shown a suspect.

4. Post-Identification Suggestive Feedback

Information received by witnesses after an identification can affect their memory.[52] Experiments conducted in the 1970's demonstrated that memories can be altered by pre-identification remarks. A study by Dr. Elizabeth Loftus involved a video that included no image of a barn. *Henderson*, 27 A.3d at 899. After viewing the video, some subjects were asked how fast the car was going when it passed a barn, and others were asked a question that included no mention of the imaginary barn. One

---

[50] Malpass & Devine, *Eyewitness Identification: Lineup Instructions and the Absence of the Offender*, 66 J. Applied Psychol. 482, 485 (1981).

[51] *See* Wright & Skagerberg, *Postidentification Feedback Affects Real Eyewitnesses*, 18 Psychol. Sci. 172, 175 (2007) (inaccuracy rate of 26.6 percent).

[52] *See, e.g.,* Bradfield et al., *The Damaging Effect of Confirming Feedback on the Relation Between Eyewitness Certainty and Identification Accuracy*, 87 J. Applied Psychol. 112 (2002).

week later, approximately 17 percent of the subjects who were asked the barn question recalled seeing it, while fewer than 3 percent in the other group had a false memory of a barn.[53]

Significantly, a witness's confidence in his or her observations may increase with information received during the investigation, a phenomenon referred to as "confidence malleability."[54] As *Henderson* found, confirmatory or post-identification feedback, which occurs when police signal to eyewitnesses that they correctly identified the suspect, can reduce doubt and engender a false sense of confidence in a witness. 27 A.3d at 899. The substantial research about confirmatory feedback includes a meta-analysis of 20 studies encompassing 2,400 identifications that found that witnesses who received feedback "expressed significantly more . . . confidence in their decision compared with participants who received no feedback."[55] This 2006 meta-analysis supports the simple point that eyewitnesses who were told that they had made a correct identification were more certain than before that they had correctly identified the target.

The more striking, and certainly more troubling, finding in this extensive research is that the subjects given confirmatory feedback recon-

---

[53] Loftus, *Leading Questions and the Eyewitness Report*, 7 Cognitive Psychol. 560, 566 (1975).
[54] *See* Wright & Skagerberg, *supra* note 51.
[55] *See* Douglass & Steblay, *Memory Distortion in Eyewitnesses: A Meta-Analysis of the Post-Identification Feedback Effect*, 20 Applied Cognitive Psychol. 859 (2006).

structed their memories of the event in such a way that in their revised memories they had a better opportunity to view the perpetrator than in reality. Scientific research shows that "those who receive a simple post-identification confirmation regarding the accuracy of their identification significantly inflate their reports to suggest better witnessing conditions at the time of the crime, stronger memory at the time of the lineup, and sharper memory abilities in general."[56] In one study, involving an analysis of identifications from line-ups by eyewitnesses to actual crimes,[57] the eyewitnesses were asked to rate three meta-memory aspects of their ability to make an accurate identification on a scale of one to ten and then were told whether they had identified the suspect or a filler. The study found that witnesses who were told they had correctly identified the suspect increased their ratings of how closely they were paying attention, how good their view was and how easily they had identified the subject. Conversely, witnesses who were told they had not identified the suspect were more likely to report that they did not have a good view or that the task was more difficult. The counter-intuitive finding that witnesses receiving confirmatory feedback change their memory for the <u>circumstances</u> of the

---

[56] *Id.* at 864-65; *see also* Wells & Bradfield, *"Good, You Identified the Suspect": Feedback to Eyewitnesses Distorts their Reports of the Witnessing Experience*, 83 J. of Applied Psychol. 360 (1998).
[57] *See* Wright & Skagerberg, *Postidentification Feedback Affects Real Eyewitnesses*, 18 Psychol. Sci. 172 (2007).

event is demonstrated by research that is consistent, reliable, and robust with large effect sizes obtained for most dependent measures (including opportunity to view, and attention paid).[58] As *Henderson* recognized, the effects of confirmatory feedback can be lasting:[59] 27 A.3d at 900.

C. Juror Research

There is strong evidence that juries widely "over believe" the reliability of eyewitness identifications, and that eyewitness testimony is therefore highly incriminating. In a pioneering demonstration of juries' credulity of eyewitnesses, Dr. Elizabeth Loftus had participants in a study read a summary of a court case that either included or did not include the positive identification testimony of an eyewitness together with some other incriminating testimony. The percentage of participants voting guilty was 18 percent when no eyewitness was present and 72 percent when one was.[60] Other studies, conducted in a variety of contexts, confirm that eyewitness testimony is persuasive.[61]

In a seminal 1983 study, researchers presented individuals with crime scenarios derived from previous empirical studies and asked the individuals to predict the accuracy rate of eyewitness identifications ob-

---

[58] *Id.*

[59] *See* Neuschatz et al., *The Effects of Post-Identification Feedback and Age on Retrospective Eyewitness Memory*, 19 Applied Cognitive Psychol. 435, 449 (2005).

[60] Loftus, *The Incredible Eyewitness*, 8(7) Psychology Today 116 (Dec. 1974).

served in the studies.[62] Nearly 84 percent of respondents overestimated the accuracy rates of identifications.[63] Moreover, the magnitude of the overestimation was significant. For example, the study's respondents estimated an average accuracy rate of 71 percent for a highly unreliable scenario in which only 12.5 percent of eyewitnesses had in fact made a correct identification.[64]

The reasons that jurors "over believe" eyewitness identifications lie in the complexity of the science and widespread lack of awareness for some of its basic precepts, as this Court recognized in *Cheatam*, ("[L]ay people may not be as aware of the malleability of eyewitness confidence or weapon focus.") 150 Wn.2d at 648, 81 P.3d at 841, citing Brian L. Cutler, *Strategies for Mitigating the Impact of Eyewitness Experts*, 37 Pros. 14, 19-20 (2003)).

D. Jury Understanding of Eyewitness Factors

Extensive surveys of lay understanding of eyewitness issues show a "discrepancy between lay understanding of factors affecting eyewitness accuracy and what decades of empirical research has reliably demonstrated to be true" and that "jurors … exhibit important limitations in their

---

[61] Leippe & Eisenstadt, *The Influence of Eyewitness Expert Testimony on Jurors' Beliefs and Judgments*, 169, 171 (Brian L. Cutler ed., 2009).
[62] *See* Brigham & Bothwell, *The Ability of Prospective Jurors To Estimate the Accuracy of Eyewitness Identifications*, 7 L. & Hum. Behav. 19, 22-24 (1983).
[63] *See id.* at 28.

knowledge of eyewitness issues, their knowledge diverges significantly from expert opinion, and it is not high in overall accuracy."[65]

Lay understanding of the various system and estimator variables and their impact on eyewitness testimony varies widely depending on the variable in question.[66] As this Court recognized in *Cheatam,* some estimator factors are "common sense," and easily understood by jurors. *Cheatam,* 150 Wn.2d at 648. A meta-analysis of surveys assessing lay knowledge of eyewitness issues testing 16 variables found that lay witnesses did not understand 75 percent of them, including cross-race bias and weapons focus.[67]

This meta-analysis, completed in 2010, reviewed 23 surveys, with a total of 4,669 participants, performed over the past 30 years.[68] Accurate answers by lay persons were recorded only 67 percent of the time, with performance for estimator variables not as strong as for system variables.[69] Notably, estimator variables that are more abstract and harder to quantify,

---

[64] *See id.* at 24.

[65] Benton et al., *Eyewitness Memory is Still Not Common Sense: Comparing Jurors, Judges and Law Enforcement to Eyewitness Experts,* 20 Applied Cognitive Psychol. 115, 126 (2006).

[66] *See, e.g.,* Desmarais & Read, *After 30 Years, What Do We Know about What Jurors Know? A Meta-Analytic Review of Lay Knowledge Regarding Eyewitness Factors,* 30 L. & Hum. Behav. 200 (2010).

[67] *Id.* at 203.

[68] *Id.* at 202.

[69] *Id.* at 205. A 2006 study concluded the same, observing "It is possible that the impact which system variables have on eyewitness accuracy may be easier to assess or assimi-

24

such as own-race bias and weapon focus, are demonstrably less under-

stood by jurors than other variables, such as the opportunity to view the

perpetrator's face and alcohol intoxication.

Extensive studies of jury understanding of own-race bias demon-

strate that jurors do not understand the impact of own-race bias on identi-

fications.[70] The 2010 meta-analysis of 23 surveys (all of which looked at

juror awareness of own-race bias) found an over-all accurate response rate

of only 57 percent.[71] Own-race bias is thus considered "beyond the ken"

of potential jurors.[72]

III.   THE RISK OF AN ERRONEOUS IDENTIFICATION IN THIS CASE

The defendant, an African-American male, was identified by a

white male following an incident that occurred at dusk and allegedly in-

volved a weapon. The defendant is a different height and weight than the

man in the victim's vague initial description. The perpetrator was report-

edly wearing a baseball cap and sunglasses at the time of the incident,

concealing his eyes. In short, the different races of the two men, the "dis-

---

late, simply because lay people have fewer existing beliefs that are contradictory and well-entrenched about the nature of their influence." Benton et al., *supra* note 73, at 126.

[70] Schmechel et al., *Beyond the Ken? Testing Jurors' Understanding of Eyewitness Reliability Evidence*, 46 Jurimetrics 177, 200, 204 (2006) (noting that almost half of the study's survey respondents thought cross-race identifications were of the same reliability as same-race identifications).

[71] Desmarais & Read, *supra* note 74, at 203.

guise" worn by the perpetrator, the presence of a weapon, and the stress of the confrontation are estimator variables that can separately and in combination reduce the likelihood of an accurate identification.

The identification procedure used here further increased the risk of an erroneous identification. The defendant was detained based on his proximity to the altercation, his clothing, and a vague physical description, including his race. The eyewitness was informed that the police had stopped a suspect and was asked to identify him. The eyewitness saw Mr. Allen, surrounded by police officers. Mr. Allen's face was made less visible when an officer asked him to lower his cap and cover his face with sunglasses that he was carrying. This showup was conducted under suggestive circumstances and with Mr. Allen intentionally presented to look like the described perpetrator. The resulting identification, and the subsequent in-court identification of Mr. Allen,[73] may have been affected by these system variables.

---

[72] *Id.* at 209. "Weapon focus" is also "beyond the ken." *Id.* At 52 percent, this variable was the third lowest average of the 16 variables included in the study, and was constant over time. *Id.* at 203, 206.

[73] Scientists have also researched a "commitment effect," where a witness tends to continue to make the same identification in subsequent identification procedures, even if the first identification was erroneous. Deffenbacher et al., *Mugshot Exposure Effects: Retroactive Interference, Mugshot Commitment, Source Confusion, and Unconscious Transference*, 30 L. & Hum. Behav. 287 (2006).

IV. IMPROVING CRIMINAL TRIALS WHERE EYEWITNESS IDENTIFICATION IS ADMITTED

A. Limitations of Traditional Adversarial Tools

Courts have historically relied on cross-examination as a safeguard against erroneous eyewitness identification.[74] This safeguard will be effective only if, *inter alia*, the defense attorney is familiar with the conditions that affect the accuracy of an identification and asks questions that elicit relevant information, the jurors understand the information highlighted in the cross-examination and incorporate that knowledge in their decision-making, and the eyewitness accurately provides testimony reflective of the actual event and not subsequent memory that has been influenced by feedback.[75] Research calls into question the likelihood that any of these assumptions will typically hold true.[76]

Cross-examination of an eyewitness is especially ineffective with respect to own-race bias. The value of cross-examination lies in its capacity to elicit facts known but not disclosed by the witness. Own-race bias, however, may well not be appreciated by an eyewitness, and questioning on the point would only result in a denial, rather than exposure of the impairment. Moreover, own-race bias cannot be elicited by any physical

---

[74] Devenport et al., *Effectiveness of Traditional Safeguards Against Erroneous Conviction Arising from Mistaken Eyewitness Identification* in *Expert Testimony on the Psychology of Eyewitness Identification* 51, 57 (Brian L. Cutler ed., 2009).
[75] *See generally id.*

demonstration as can other impairments, such as poor eyesight or a poor memory. In short, none of the time-honored forms of cross-examination will demonstrate own-race bias.[77]

B.   Usefulness of Expert Testimony

Multiple studies evaluating layperson knowledge of factors affecting eyewitness have shown that jurors do not understand many of the variables that influence eyewitness identifications,[78] as explained *supra*. Even when they do, they do not know how to apply their understanding to the interpretation of evidence.[79] This limited understanding by non-experts of the 30 years of scientific research can apply to attorneys and judges, as well as jurors, limiting effectiveness of traditional safeguards.[80] Expert testimony supplies a means to transfer or deliver this body of research to the courtroom. Experts in eyewitness identification can alert jurors to the existence of factors known to have effects on identification, and to indicate how strong those effects have been shown to be. In that sense, such experts are educators, rather than providers of ultimate opinion testimony (*i.e.*, on whether a specific eyewitness made an accurate identification).

---

[76] *See generally id.*
[77] Johnson, *Cross-Racial Identification Errors in Criminal Cases*, 69 Cornell L. Rev. 934 (1983).
[78] Malpass et al., *The Need for Expert Psychological Testimony on Eyewitness Identification*, in *Expert Testimony on the Psychology of Eyewitness Identification* 3, 9 (Cutler ed., 2009).
[79] *Id.*

This Court has, of course, recognized in *Cheatam* that some eyewitness-identification issues, such as cross-racial identification, weapons focus and stress, are appropriate subjects for expert testimony.[81]

Expert testimony on eyewitness issues has been shown to improve juror knowledge about influences on eyewitness reports.[82] A 1989 study by Amicus Steven Penrod and others tested a large sample of college students and experienced jurors using a videotape of an armed robbery trial and various witnessing and identification conditions.[83] Expert testimony was found to increase juror sensitivity to both types of conditions: without expert testimony, the number of guilty verdicts was the same for both good and poor conditions, but, with expert testimony, guilty verdicts were significantly higher when conditions were good than when they were poor. A later study also demonstrated that expert testimony results in fewer guilty verdicts when viewing conditions are poor,[84] suggesting that the effect is not simply skepticism, but appropriate skepticism. In another study, mock jurors rated an eyewitness identification that was flawed as

[80] Devenport et al., *Eyewitness Identification Evidence: Evaluating Commonsense Evaluations*, 3 Psychol. Pub. Pol'y & L. 338 (1997).

[81] *State v. Cheatam*, 150 W.2d 626, 648, 81 P.3d 830 (2003).

[82] Ramirez et al., *Judges' Cautionary Instructions on Eyewitness Testimony*, 14 Am. J. Forensic Psychol. 31 (1996); Devenport et al., *How Effective Are Cross-Examination and Expert Testimony Safeguards? Jurors' Perceptions of the Suggestiveness and Fairness of Biased Lineup Procedures*, 87 J. Applied Psychol. 1042 (2002).

[83] Cutler et al., *The Eyewitness, the Expert Psychologist, and the Jury*, 13 Law & Hum. Behav. 311 (1989).

more unfair and the defendant as less culpable when there was expert testimony.[85] The positive effect of expert testimony has been shown to depend on various factors, including whether jurors hear it after the eyewitness testimony and whether the judge reminded the jurors of what the expert said as part of the final instructions.[86] Research continues on the impact of expert testimony on this complex and still-developing body of research.

The reality is, however, that not all defendants can afford to retain expert witnesses, and that the demand for experts with this specialization exceeds the small supply. Therefore, while expert testimony is the most demonstrably effective means to educate juries, it is not always available.

C. Eyewitness Identification Jury Instructions

In recent years, a number of state courts have come to view comprehensive jury instructions as a cost-effective and efficient way to educate jurors about eyewitness identification issues. In a ruling issued this week, the U.S. Supreme Court said that "[e]yewitness-specific jury instructions, which many federal and state courts have adopted," are an im-

---

[84] Leippe et al., *Timing of Eyewitness Expert Testimony, Jurors' Need for Cognition, and Case Strength as Determinants of Trial Verdicts*, 89 J. Applied Soc. Psychol. 201 (2004).
[85] Devenport et al., *How Effective Are Cross-Examination and Expert Testimony Safeguards? Jurors' Perceptions of the Suggestiveness and Fairness of Biased Lineup Procedures*, 87 J. Applied Psychol. 1042 (2002).
[86] Leippe et al., *supra* not 94, at 201.

portant safeguard. *Perry v. New Hampshire*, 565 U. S. ____ (2012) (slip op., at 16).

Some courts have developed jury instructions that resemble expert psychological testimony. *See, e.g., People v. Wright*, 729 P.2d 280 (CAL. 1987). Instructions patterned on *Wright* bring to jurors' attention factors bearing on the unreliability of eyewitness identifications. They can also be customized so that they highlight the important factors in any given case. The New Jersey Supreme Court in *Henderson* has furthered this initiative by directing a task force to devise science-based, issue-specific instructions. *Henderson*, 27 A.3d at 878.[87] (To date, there is no published research on the efficacy of issue-specific instructions, including cross-race bias, a need that researchers are seeking to satisfy.)

These kinds of science-based, issue-specific jury instructions represent a sea change from earlier iterations of jury instructions in the eyewitness identification context, such as those based on *Telfaire*, supra, which list only a few factors, such as the viewing opportunity, and, more

---

[87] Several states have adopted jury instructions that specifically offer guidance on how to evaluate the credibility of a witness's testimony in light of many of the scientific findings discussed herein. These states include Alaska, Arizona, California, Connecticut, Georgia, Hawaii, Kansas, Maryland, Massachusetts, New Hampshire, New York, North Carolina, Ohio, Oklahoma, and Pennsylvania.

important, do not indicate to juries the possible effect of these factors. Thus, *Telfaire* instructions are considered problematic.[88]

There is a limited amount of research on the effectiveness of broad jury instructions on eyewitness identification issues, and the few existing studies do not show consistent results. It is known that *Telfaire* instructions do not improve juror understanding, most likely because they are not sufficiently informative.[89] On the other hand, research shows that jury instructions that are clearly worded and explain how various factors affect the reliability of an eyewitness identification improve juror understanding of eyewitnessing factors more than do *Telfaire* instructions.[90] Another study, which generally yielded conflicting results, did show that mock jurors who heard *Telfaire* instructions gave less weight to weak eyewitness testimony than mock jurors who received no instructions.[91]

More significant is the large body of research on jury instructions generally, prompted by the judiciary's assumption that juries actually pay attention to and follow jury instructions. *Richardson v. Marsh*, 481 U.S.

---

[88] Devenport *et al.*, *Effectiveness of Traditional Safeguards Against Erroneous Conviction Arising from Mistaken Eyewitness Identification, in Expert Testimony on the Psychology of Eyewitness Identification* 51, 62 (Brian L. Cutler ed., 2009).

[89] *See* Cutler et al., *Nonadversarial Methods for Improving Juror Sensitivity to Eyewitness Evidence*, 20 J. Appl. Soc. Psychol. 1197, 1198-1200, 1202-06 (1990)

[90] *See* Ramirez et al., *Judges Cautionary Instructions on Eyewitness Testimony*, 14 Am. J. Forensic Psychol. 31, 56 (1996) ("The revised instructions . . . led to a significant effect on the subjects' expert knowledge of eyewitness factors.").

[91] *See* Greene, *Judges' Instruction on Eyewitness Testimony*, 18 J. Appl. Soc. Psychol. 252, 260 (1988).

200, 211 (1987) ("The rule that juries are presumed to follow their instructions is a pragmatic one, rooted less in the absolute certitude that the presumption is true than in the belief that it represents a reasonable practical accommodation of the interests of the state and the defendant in the criminal justice process."). A good is example of these useful studies is a meta-analysis involving studies of jury instructions to ignore inadmissible evidence. The meta-analysis by amicus Dr. Nancy K. Steblay and others found that when judges provided a clear rationale for a ruling of inadmissibility, juror compliance increased.[92] Jurors, in other words, responded to specific information about why the inadmissible evidence was unreliable or had no bearing on the case. Similarly, a study involving jury instructions on substantive law showed that the timing and repetition of jury instructions can enhance juror understanding.[93] These important studies indicate that issue-specific instructions on eyewitness testimony could be very effective.

The challenge in designing experiments has been, and will continue to be, the difficulty in analyzing how the members of a jury internalize what they learn from instructions and together reach a verdict that

---

[92] Steblay et al, *The Impact on Juror Verdicts of Judicial Instruction to Disregard Inadmissible Evidence: a Meta-Analysis*, 30 L. & Hum. Behav. 469 (2006).
[93] *See* Cruse & Browne, *Reasoning in a Jury Trial: The Influence of Instructions*, 114 J. Gen. Psychol. 129, 133 (1986) ("Multiple exposures to clear legal definitions serve to

reflects their instructed understanding. Early studies used fabricated eye-witness testimony, reflecting good and bad viewing conditions, and compared the verdicts of "juries" who received instructions with those which did not. The shortcoming of this approach is that a lower number of verdicts might reflect either that the "jurors" understood and correctly applied the relevant concepts (which would be appropriate) or that they were merely skeptical, and discounted the eyewitness testimony entirely.

While psychological research has not consistently found sensitization effects for general eyewitness jury instructions, research is needed to understand the effectiveness of issues-specific instructions. Research will add to the field's understanding of how jurors process and apply information about eyewitness accuracy to evaluate eyewitness evidence in eyewitness cases, will add to the body of psychological literature on safeguards against erroneous identification evidence and will provide guidance about the desirability and construction of issue-specific jury instructions. Researchers working in this field can provide the courts with valuable information about whether issue- specific eyewitness instructions are effective and if so, the method and content of issue-specific instructions that will be most effective in sensitizing jurors to eyewitness evidence. Proper methods of sensitizing jurors to eyewitness evidence have become imperative

---

equate rules among jurors and eliminate the necessity of relying on naïve assumptions to

in light of emerging statistics about the high percentage of eyewitness mis-identifications contributing to wrongful convictions, as well as the recognition by the U.S. Supreme Court in *Perry* that jury instructions provide an important safeguard against erroneous eyewitness identifications. *See* 565 U. S. ____ (2012) (slip op., at 15-16).

CONCLUSION

For the foregoing reasons, the Amici urge this Court to consider the persuasive scientific research regarding the reliability of eyewitness identifications and allow jury instructions to supplement traditional safeguards preventing convictions based on erroneous eyewitness identifications, such as expert testimony.

Dated this 13th day of January, 2012.

Respectfully submitted,

/s/ *Todd Maybrown*
Attorney for *Amici Curiae*
College and University Professors

---

evaluate testimony.").

Supreme Court No. 86119-6
161 Wn. App. 727, 255 P.3d 784 (May 9, 2011)

IN THE SUPREME COURT OF THE STATE OF WASHINGTON

STATE OF WASHINGTON,

Respondent,

v.

BRYAN EDWARD ALLEN,

Petitioner.

ON REVIEW FROM THE COURT OF APPEALS, DIVISION ONE

APPENDIX TO THE BRIEF OF THE *AMICI CURIAE*
COLLEGE AND UNIVERSITY PROFESSORS
JENNIFER DEVENPORT, JENNIFER E. DYSART,
GEOFFREY LOFTUS, STEVEN D. PENROD,
NANCY K. STEBLAY and GARY L. WELLS
IN SUPPORT OF PETITIONER BRYAN EDWARD ALLEN

ALLEN, HANSEN & MAYBROWN, P.S.
One Union Square
600 University Street, Suite 3020
Seattle, Washington 98101
Telephone: (206) 447-9681

MILBANK, TWEED, HADLEY & McCLOY LLP
1 Chase Manhattan Plaza
New York, New York 10005
Telephone: 212-530-5088
Of Counsel to Attorney for *Amici Curiae*
College and University Professors

Index to Appendix

The articles included in the appendix are:

Cross-Racial Facial Identification:
A Social Cognitive Integration

Tara Anthony
Carolyn Copper
Brian Mullen
Syracuse University

A meta-analytic integration of the cross-racial facial identification effect is reported. The results indicate a significant, weak-to-moderate tendency for individuals to be more accurate in the recognition of faces of in-group members than those of out-group members. There is a trend for this effect to be stronger among White subjects than Black subjects. In addition, predictors exhibited different patterns for Black subjects and White subjects. Specifically, greater depth of processing engaged by experimental instructions in these studies strengthened the effect for White subjects and weakened it for Black subjects. Similarly, the cross-racial facial identification effect increased as a function of the duration of exposure to the target faces for White subjects but decreased as a function of duration of exposure for Black subjects. These results are explained in terms of recent theoretical developments concerning intergroup phenomena, particularly in terms of mechanisms of cognitive representations of in-groups and out-groups.

Differential recognition of own- versus other-race faces has been a topic of social psychological research for over 20 years (Malpass & Kravitz, 1969). The general tendency is for subjects to exhibit superior memory for faces belonging to members of their own ethnic or racial group than for faces belonging to members of another group (Bothwell, Brigham, & Malpass, 1989). It is somewhat surprising that previous reviews and summaries of this phenomenon are generally at a loss to provide a compelling theoretical account for the basic effect. Plausible accounts, such as greater actual similarity among faces in one race than in another or differential contact or experience with the racial in-group and with racial out-groups, have generally not been supported by the evidence (see Bothwell et al., 1989; Brigham & Malpass, 1985).

It is seldom recognized that the cross-racial facial identification effect bears a striking resemblance to another phenomenon of intergroup perception: the relative heterogeneity effect (Jones, Wood, & Quattrone, 1981; Mullen & Hu, 1989). This is the general tendency for people to perceive the in-group as more heterogeneous than the out-group. Both the cross-racial facial identification effect and the relative heterogeneity effect appear to be based on a general tendency to fail to distinguish among individuals in the out-group. Brigham (1991) recently commented on the resemblance between these two phenomena of intergroup perceptions.

Such a tendency to fail to distinguish among individuals in a group has been explained in terms of the links between group composition, salience, and cognitive representations (Mullen, 1991; Mullen, Johnson, & Anthony, 1990). Specifically, as the size of the group decreases, the salience of that group increases. This increased salience leads to processing of information about that group in a prototype representation mode. Alternatively, as the size of the group increases, the salience of that group decreases. This decreased salience leads to processing of information about that group in an exemplar representation mode. According to prototype models of category

representations (Posner & Keele, 1970; Reed, 1972), a category is represented by the prototype, some "average" or most typical member. According to exemplar models of category representation (Brooks, 1978; Medin & Schaffer, 1978), a category is represented by an accumulation of information about known exemplars of that category.[1] One consequence of the prototype processing mode is a minimization of differences within the category, whereas the exemplar processing mode accommodates the handling of discrete targets with all their differences. This explains why relative heterogeneity occurs when the in-group is proportionately larger whereas relative homogeneity occurs when the in-group is proportionately smaller (see Mullen, 1991; Mullen & Hu, 1989).

Given the parallel between the relative heterogeneity effect and the cross-racial facial identification bias, this theoretical approach to social cognition and group processes may also be useful in explaining the cross-racial facial identification effect. The assumption underlying an application of this approach would be that the typical cross-racial facial identification effect stems from a simplified, prototype representation of target faces of out-group members. If this assumption is true, then factors that influence the cognitive representations used to process target faces of in-group members and out-group members should influence the cross-racial facial identification effect.

First, consider the effects of subject race on the basic cross-racial facial identification effect. The general effect would be expected to be stronger for White subjects: White subjects may be more likely to process the smaller Black out-group in a prototype representation mode, leading to a failure to discriminate among out-group exemplars. Alternatively, the general effect would be expected to be weaker for Black subjects: Black subjects may be more likely to process the larger White out-group in an exemplar processing mode, leading to more successful discrimination among out-group exemplars.

Second, consider the effects of depth of processing. Generally speaking, instructions emphasizing impression formation engage deeper, or more elaborative, cognitive processing than instructions emphasizing memorization (e.g., Hamilton, Katz, & Leirer, 1980; Pryor, 1986). In the service of forming an impression, the subjects may more effortfully engage the cognitive representations that they are inclined to use. For Black subjects, the more vigorous use of exemplar representations will increase their discrimination among out-group members, thus reducing the overall cross-racial facial identification effect. For White subjects, the more vigorous use of prototype representations will decrease their discrim-

ination among out-group members, thus enhancing the overall cross-racial facial identification effect.

Third, consider the effects of duration of initial exposure to the target faces. Generally speaking, the longer one can examine the target faces, the better one's retrieval of those faces ought to be, assuming that one is processing the information about the target faces with an eye for their individuating differences. This should characterize the Black subjects in their more exemplar representation processing of the larger White out-group, and so Black subjects should show a lesser in-group identification bias with longer exposure to target faces. However, White subjects are less likely to be processing information about the target faces with an eye for their individuating differences. Rather, White subjects engaging in more prototype processing of the smaller Black out-group should show an even greater in-group identification bias with longer exposure to target faces.

This article reports the results of a meta-analysis (Mullen, 1989; Mullen & Rosenthal, 1985; Rosenthal, 1984) integrating the research examining the cross-racial facial identification effect. The goals of this integration were to compare the cross-racial facial identification effect obtained for White and Black subjects and to gauge the effects of depth of processing and duration of exposure to target faces for White and for Black subjects.

METHOD

In accord with the procedures specified in Mullen (1989) and Rosenthal (1984), an exhaustive manual and computer search of the literature was conducted. Studies were selected for inclusion if they reported (or intelligibly implied) some comparison of the identification of White and Black targets for White subjects and some comparison of the identification of White and Black targets for Black subjects. When necessary, the original authors were contacted and supplementary information was obtained. This resulted in a total of 15 studies, with 22 separate hypothesis tests for White subjects and a matched set of 22 hypothesis tests for Black subjects, representing the responses of a total of 1,725 subjects. The hypothesis tests included in this meta-analysis, along with the relevant statistical information, are presented in Table 1.

In addition to the basic statistical information (statistical test of the hypothesis, corresponding degrees of freedom, sample size, and direction of effect) and the race of the subjects, two predictors were derived for each hypothesis test. The depth of processing engaged by the experimental procedure was rated for those studies that reported enough detail regarding instructions delivered

PERSONALITY AND SOCIAL PSYCHOLOGY BULLETIN

TABLE 1: Studies Included in the Meta-analysis

Study	Statistic	z	r	N	DOE[a]	DUR[b]	DOP[c]	Race[d]
Barkowitz & Brigham (1982), Study 1	$t(213) = 6.670$	6.639	.416	52	+	1.5	−0.02	W
	$t(213) = 5.589$	5.393	.358	52	+	1.5	−0.02	W
	$t(213) = 2.344$	2.326	.159	52	+	1.5	−0.02	W
	$t(213) = 0.650$	−0.649	−.044	27	−	1.5	−0.02	B
	$t(213) = 1.429$	−1.424	−.097	27	−	1.5	−0.02	B
	$t(213) = 0.455$	−0.454	−.31	27	−	1.5	−0.02	B
Barkowitz & Brigham (1982), Study 2	$t(93) = 2.379$	2.338	.240	43	+	1.5	−0.02	W
	$t(93) = 3.361$	3.262	.330	58	+	1.5	−0.02	B
Brigham & Barkowitz (1978)	$t(154) = 5.432$	5.187	.401	76	+	1.8	—	W
	$t(154) = 8.887$	7.972	.582	86	+	1.8	—	B
Brigham & Williamson (1979)	$t(39) = 5.021$	4.381	.627	27	+	1.8	−0.02	B
	$t(39) = 0.634$	0.628	.101	14	+	1.8	−0.02	W
Chance, Goldstein, & McBride (1975), Study 2	$t(88) = 3.361$	3.251	.337	48	+	3.0	−1.47	W
	$t(88) = 2.922$	2.847	.297	48	+	3.0	−1.47	B
Cross, Cross, & Daly (1971)	$r(10) = .682$	2.438	.682	150	+	—	2.92	W
	$r(10) = .329$	−1.04	−.329	150	−	—	2.92	B
Devine & Malpass (1985)	$t(42) = 1.479$	1.452	.222	8	+	3.0	2.62	B
	$t(42) = 1.601$	1.568	.240	8	+	3.0	2.62	W
	$t(42) = 4.487$	4.031	.569	8	+	3.0	2.32	B
	$t(42) = 3.226$	3.031	.446	8	+	3.0	2.32	W
	$t(42) = 2.959$	2.803	.415	8	+	3.0	−1.77	B
	$t(42) = 2.037$	1.977	.300	8	+	3.0	−1.77	W
Ellis & Deregowski (1981)	$t(92) = 0.894$	0.890	.093	96	+	5.0	−2.66	W
	$t(92) = 2.683$	2.625	.269	96	+	5.0	−2.66	B
Feinman & Entwisle (1976)	$t(256) = 6.168$	5.950	.630	96	+	4.0	—	W
	$t(256) = 5.142$	5.011	.306	96	+	4.0	—	B
	$t(256) = 3.200$	3.166	.196	48	+	4.0	—	W
	$t(256) = 2.094$	2.083	.130	48	+	4.0	—	B
Galper (1973)	$t(13) = 1.530$	−1.439	−.391	14	−	9.4	−1.48	W
	$t(13) = 1.859$	1.717	.458	14	+	9.4	−1.48	B
	$t(13) = 3.293$	2.754	.674	14	+	9.4	−1.48	W
	$t(15) = 3.724$	3.081	.693	16	+	9.4	−1.48	B
Lindsay, Jack, & Christian (1990)	$t(30) = 0.777$	−0.767	−.140	16	−	0.12	−2.08	B
	$t(30) = 4.994$	4.225	.674	16	+	0.12	−2.08	W
Malpass (1974)	$t(60) = 10.254$	7.762	.798	14	+	1.1	—	B
	$t(60) = 8.554$	−6.888	−.741	18	−	1.1	—	W
Malpass & Kravitz (1969)	$t(24) = 0.294$	−0.291	−.060	13	−	1.5	1.45	B
	$t(24) = 3.582$	3.172	.590	13	+	1.5	1.45	W
	$t(12) = 0.201$	−0.197	−.058	7	−	1.5	1.45	B
	$t(12) = 2.518$	2.209	.588	7	+	1.5	1.45	W
Malpass, Lavigueur, & Weldon (1973), Study 1	$t(112) = 1.980$	1.959	.184	16	+	1.0	—	B
	$t(112) = 1.890$	1.871	.176	18	+	1.0	—	W
Sheperd, Deregowski, & Ellis (1974)	$t(60) = 2.278$	2.222	.282	32	+	4.0	0.28	B
	$t(60) = 6.007$	5.293	.613	32	+	4.0	0.28	W

a. DOE = direction of effect. Positive direction of effect indicates greater memory for in-group faces than for out-group faces. Negative direction of effect indicates greater memory for out-group faces than in-group faces.
b. DUR = duration of exposure to target faces, in seconds per face.
c. DOP = depth of processing. Larger and positive numbers indicate more impression formation and less memorization.
d. Race: B = Black subjects, W = White subjects.

to subjects. Three judges read the experimental instructions given in each study and rated the amount of impression formation and the amount of memorization generated by the instructions. The mean interjudge reliability was $\bar{r} = .213$ for impression formation and $\bar{r} = .598$ for memorization. These rendered a Spearman-Brown reliability for $R = .448$ for impression formation and $R = .817$ for memorization. Conceptually, impression formation and memorization should be inversely related, and indeed the mean judges' ratings for these two attributes correlated $r = -.397$, $p = .01005$. Therefore, a single, global index of depth of processing was derived by standardizing these judges' ratings, reverse-scoring the standardized ratings for memorization, and then averaging these two standard ratings. This composite rating for depth of processing increased as impression formation increased and memorization decreased.[2] *Duration of exposure* to the target stimuli was simply coded from the procedure section of each study.

RESULTS

General effects. Overall, there was a significant, $z = 13.707$, $p = 5.21E-31$, weak-to-moderate, Fisher's $\bar{z} = 0.292$, $\bar{r} = .284$, $\bar{r}^2 = .081$, tendency to remember in-group faces better than out-group faces. This overall effect is slightly lower than the $\bar{r} = .330$, $\bar{r}^2 = .109$ reported by Bothwell et al. (1989).

Effects of subject race. The cross-racial facial identification effect was significant, $z = 10.638$, $p = 9.27E-23$, and moderate, Fisher's $\bar{z} = 0.359$, $\bar{r} = .345$, $\bar{r}^2 = .119$, for White subjects. This effect was significant, $z = 8.727$, $p = 3.12E-17$, but weak, Fisher's $\bar{z} = 0.221$, $\bar{r} = .218$, $\bar{r}^2 = .048$, for Black subjects. The difference between these two effects was not significant, $z = 0.510$, $p = .3051$. Thus, the cross-racial facial identification effect was stronger among White subjects than among Black subjects, but this trend did not achieve significance.

Effects of elaborative processing. A significant interaction between subject race and elaborative processing was observed, $z = 1.833$, $p = .0334$. Among White subjects, the cross-racial identification effect increased as a function of elaborative processing, $r = .266$, $z = 1.078$, $p = .1404$. Among Black subjects, the cross-racial facial identification effect decreased as a function of elaborative processing, $r = -.350$, $z = 1.514$, $p = .0650$.

Effects of duration of exposure. A significant interaction between subject race and duration of exposure was observed, $z = 2.419$, $p = .00777$. Among White subjects, the longer the duration of the initial exposure to the target faces, the stronger the cross-racial identification effect,

$r = .200$, $z = 2.142$, $p = .0161$. Among Black subjects, the longer the duration of the initial exposure to the target faces, the weaker the cross-racial facial identification effect, $r = -.135$, $z = 1.276$, $p = .1010$.

DISCUSSION

These results reveal a significant, weak-to-moderate tendency for people to exhibit better memory for faces of in-group members than for faces of out-group members. There was a nonsignificant trend for this effect to be stronger for White subjects: The cross-racial facial identification bias accounted for 2.5 times the variance in recognition for White subjects ($\bar{r} = .345$, $\bar{r}^2 = .119$) as that for Black subjects ($\bar{r} = .218$, $\bar{r}^2 = .048$). This trend is different from the effects reported in the Bothwell et al. (1989) meta-analysis, which demonstrated nearly identical cross-racial facial identification for Black subjects ($\bar{r} = .335$, $\bar{r}^2 = .112$) and White subjects ($\bar{r} = .326$, $\bar{r}^2 = .106$). This difference might be attributed, in part, to different meta-analytic databases, insofar as the present effort included 16 hypothesis tests not included in Bothwell et al. (1989).

The tendency for the cross-racial facial identification effect to be stronger among White subjects than among Black subjects is consistent with the reasoning developed above. To the extent that the basic cross-racial facial identification effect represents prototype processing of information about the out-group, this effect should be stronger among the members of the larger White group, who will already be processing information about the smaller Black out-group in prototype representation mode. Although this is precisely the pattern that occurred, this tendency was not significant and should be interpreted with due caution.

The contrasting effects of elaborative processing also make sense within the reasoning developed above. Black subjects are apparently processing information about targets in the larger White out-group in an exemplar representation mode. Black subjects will be even more likely to engage these "bottom-up," discrimination-enhancing cognitive representations when instructions promote more elaborative processing and thus will make fewer errors. Alternatively, White subjects are apparently processing information about targets in the smaller Black out-group in a prototype processing mode. White subjects will be even more likely to engage these "top-down," discrimination-reducing cognitive representations when instructions promote more elaborative processing and thus will make more errors.

The contrasting effects of duration of exposure to the target faces also make sense within the rationale pre-

PERSONALITY AND SOCIAL PSYCHOLOGY BULLETIN

sented above. With longer duration of exposure, Black subjects who are processing information about targets in the larger White out-group in an exemplar representation mode will be better able to take advantage of individuating and distinguishing features in those target faces. Alternatively, with longer duration of exposure, White subjects who are processing information about targets in the smaller Black out-group in a prototype processing mode seem even more likely to distort and lose any individuating and distinguishing features in those target faces.

The results presented here lend tentative support to a social cognitive explanation for the cross-racial facial identification effect. Similar to the relative heterogeneity effect, the cross-racial facial identification effect may have its roots in basic cognitive mechanisms, operating at a level that is at once more fundamental and more subtle than the more macro explanations considered for this phenomenon in the past (e.g., degree of experience with the racial out-group; veridical differences in the discriminability of faces in certain racial groups).

One interesting implication of this conclusion is that the cross-racial facial identification effect is simply a special case of a more general phenomenon. The cross-racial facial identification effect, like the relative heterogeneity effect, may represent a specific operationalization of the general tendency to fail to distinguish among members of the out-group. At a more specific level, a more provocative implication is that there may not be a cross-racial facial identification effect per se. Rather, there may be a cross-*group* facial identification effect: Subjects may exhibit superior memory for faces belonging to members of their own group, regardless of the basis for that group distinction. This could explain why attempts to account for the cross-racial facial identification effect with explanations and mechanisms directly linked to race have been so unsuccessful.

Although the forensic concerns that have driven this research have been useful in highlighting the practical importance of the cross-racial facial identification effect, it may well be that a more basic, theoretical approach will be needed to progress to functional explanations for this phenomenon. The present patterns suggest that the failure to distinguish among faces of out-group members may be driven by the same types of basic cognitive mechanisms that drive other in-group/out-group social cognitive phenomena, such as the relative heterogeneity effect (Mullen & Hu, 1989), social projection (Mullen & Hu, 1988), and in-group bias (Mullen, Brown, & Smith, 1990). Future studies of cross-racial facial identification should begin to examine these basic cognitive mechanisms.

NOTES

1. Although prototype and exemplar models have in the past been treated as mutually exclusive, recent evidence indicates that both prototype representations and exemplar representations can contribute to the handling of information about category members, both for nonsocial information tasks (Medin, Altom, & Murphy, 1984; Smith & Medin, 1981) and for social information tasks (Linville, Fischer, & Salovey, 1989; Park & Hastie, 1987). A considerable amount of research (e.g., Medin et al., 1984; Sherman & Corty, 1984; Smith & Zarate, 1990) converges on the tendency for people to use prototypes when those prototypes are primed or made easily accessible.

2. We would like to thank Dave Schneider for helpful comments about the interrelations between impression formation, memorization, and depth of processing. It should be noted that the results reported for this composite index of depth of processing were replicated with the two separate component ratings of impression formation and memorization (e.g., White subjects exhibited stronger cross-racial facial identification effects when instructions emphasized impression formation and weaker effects when instructions emphasized memorization, whereas Black subjects exhibited the opposite pattern).

REFERENCES

Barkowitz, P., & Brigham, J. C. (1982). Recognition of faces: Own-race bias, incentive, and time delay. *Journal of Applied Social Psychology, 12,* 255-268.
Bothwell, R. K., Brigham, J. C., & Malpass, R. S. (1989). Cross-racial identification. *Personality and Social Psychology Bulletin, 15,* 19-25.
Brigham, J. C. (1991). *Social psychology.* New York: Harper Collins.
Brigham, J. C., & Barkowitz, P. (1978). Do "they all look alike"? The effect of race, sex, experience, and attitudes on the ability to recognize faces. *Journal of Applied Social Psychology, 8,* 306-318.
Brigham, J. C., & Malpass, R. S. (1985). The role of experience and contact in the recognition of faces of own- and other-race persons. *Journal of Social Issues, 41,* 139-155.
Brigham, J. C., & Williamson, N. L. (1979). Cross-racial recognition and age: When you're over 60, do they still "all look alike"? *Personality and Social Psychology Bulletin, 5,* 218-222.
Brooks, L. (1978). Nonanalytic concept formation and memory for instances. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization.* Hillsdale, NJ: Lawrence Erlbaum.
Chance, J., Goldstein, A. G., & McBride, L. (1975). Differential experience and recognition memory for faces. *Journal of Social Psychology, 97,* 243-253.
Cross, J. F., Cross, J., & Daly, J. (1971). Sex, race, age, and beauty as factors in recognition of faces. *Perception and Psychophysics, 10,* 393-396.
Devine, P. G., & Malpass, R. S. (1985). Orienting strategies in differential face recognition. *Personality and Social Psychology Bulletin, 11,* 33-40.
Ellis, H. D., & Deregowski, J. B. (1981). Within-race and between-race recognition of transformed and untransformed faces. *American Journal of Psychology, 94,* 27-35.
Feinman, S., & Entwisle, D. R. (1976). Children's ability to recognize other children's faces. *Child Development, 47,* 506-510.
Galper, R. E. (1973). "Functional race membership" and recognition of faces. *Perceptual and Motor Skills, 37,* 455-462.
Hamilton, D. L., Katz, L. B., & Leirer, V. O. (1980). Organizational processes in impression formation. In R. Hastie, T. Ostrom, E. Ebbesen, R. Wyer, Jr., D. Hamilton, & D. Carlston (Eds.), *Person memory: The cognitive basis of social perception.* Hillsdale, NJ: Lawrence Erlbaum.
Jones, E. E., Wood, G. C., & Quattrone, G. A. (1981). Perceived variability of personal characteristics in ingroups and outgroups: The role of knowledge and evaluation. *Personality and Social Psychology Bulletin, 7,* 523-528.
Lindsay, D. S., Jack, P. C., & Christian, M. (1990). *Other-race face perception.* Unpublished manuscript, McMaster University.
Linville, P., Fischer, G. W., & Salovey, P. (1989). Perceived distributions of the characteristics of ingroup members and outgroup members.

Journal of Personality and Social Psychology, 57, 165-188.

Malpass, R. S. (1974). Racial bias in eyewitness identification. *Personality and Social Psychology Bulletin, 1,* 42-44.

Malpass, R. S., & Kravitz, J. (1969). Recognition for faces of own and other race. *Journal of Personality and Social Psychology, 13,* 330-334.

Malpass, R. S., Lavigueur, H., & Weldon, D. E. (1973). Verbal and visual training in face recognition. *Perception and Psychophysics, 14,* 285-292.

Medin, D. L., Altom, M. W., & Murphy, T. D. (1984). Given versus induced category representations: Use of prototype and exemplar information in classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 333-352.

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review, 85,* 207-238.

Mullen, B. (1989). *Advanced BASIC meta-analysis.* Hillsdale, NJ: Lawrence Erlbaum.

Mullen, B. (1991). Group composition, salience, and cognitive representations: The phenomenology of being in a group. *Journal of Experimental Social Psychology, 27,* 1-27.

Mullen, B., Brown, R., & Smith, C. (1990, August). *Ingroup bias as a function of salience: The effects of proportionate ingroup size and the reality of the groups.* Paper presented at the 98th annual meeting of the American Psychological Association, Boston.

Mullen, B., & Hu, L. (1988). Social projection as a function of cognitive mechanisms: Two meta-analytic integrations. *British Journal of Social Psychology, 27,* 333-356.

Mullen, B., & Hu, L. (1989). Perceptions of ingroup and outgroup variability: A meta-analytic integration. *Basic and Applied Social Psychology, 10,* 233-252.

Mullen, B., Johnson, C., & Anthony, T. (1990). *Prototype vs. exemplar processing of ingroup versus outgroup as a function of relative group size: The phenomenology of being in a group.* Unpublished manuscript, Syracuse University.

Mullen, B., & Rosenthal, R. (1985). *BASIC meta-analysis: Procedures and programs.* Hillsdale, NJ: Lawrence Erlbaum.

Park, B., & Hastie, R. (1987). Perception of variability in category development: Instance versus abstraction-based stereotypes. *Journal of Personality and Social Psychology, 53,* 621-635.

Posner, M. I., & Keele, S. W. (1970). Retention of abstract ideas. *Journal of Experimental Psychology, 83,* 304-308.

Pryor, J. B. (1986). The influence of different encoding sets upon the formation of illusory correlations and group impressions. *Personality and Social Psychology Bulletin, 12,* 216-226.

Reed, S. K. (1972). Pattern recognition and categorization. *Cognitive Psychology, 3,* 382-407.

Rosenthal, R. (1984). *Meta-analytic procedures for social research.* Beverly Hills, CA: Sage.

Sheperd, J. W., Deregowski, J. B., & Ellis, H. D. (1974). A cross-cultural study for recognition memory for faces. *International Journal of Psychology, 9,* 205-211.

Sherman, S. J., & Corty, E. (1984). Cognitive heuristics. In R. S. Wyer & T. K. Srull (Eds.), *Handbook of social cognition.* Hillsdale, NJ: Lawrence Erlbaum.

Smith, E. E., & Medin, D. L. (1981). *Categories and concepts.* Cambridge, MA: Harvard University Press.

Smith, E. R., & Zarate, M. A. (1990). Exemplar and prototype use in social cognition. *Social Cognition, 8,* 243-262.

Meta-Analysis

Cross-Racial Identification

Robert K. Bothwell
University of Texas at El Paso
John C. Brigham
Florida State University
Roy S. Malpass
State University of New York at Plattsburgh

1180

This article reviews the research on differential recognition for own- versus other-race faces. A meta-analysis of 14 samples revealed that the magnitude of the own-race bias is similar for both Black and White subjects, accounting for about 10% of the variance in recognition accuracy. There is a considerable consistency across studies, indicating that memory for own-race faces is superior to memory for other-race faces. Both Black and White subjects exhibited own-race bias in 79% of the samples reviewed.

Over the last decade the literature on psychological factors affecting eyewitness identifications has grown dramatically in size. It has also grown in the demands placed upon it, because theorists have stressed the importance of research that is directly applicable to police lineup techniques and courtroom procedures. An important part of this process is the reexamination of earlier literature, because the field has advanced and the focus of research questions has sharpened.

One such area of reexamination is that of the differential recognition between American Blacks and Whites for faces of their own and of the other race. Legal scholars have expressed concern over an *own-race bias* in eyewitness identifications for quite some time. Feingold (1914, p. 50) asserted that it is "well known that, other things being equal, individuals of a given race are distinguishable from each other in proportion to our familiarity, to our contact with the race as a whole. Thus, to the uninitiated American, all Asiatics look alike, while to the Asiatic all White men look alike."

Most experts in the field of eyewitness memory and about half of potential jurors endorse the belief that cross-racial identifications are less reliable than same-race identifications (Yarmey & Jones, 1983). This presumption is based on the belief in the existence of an own-race bias: that people recognize people of their own race better than people of another race. At the same time, Lindsay and

---

Wells (1983) have questioned whether the effect exists at all. Clearly the evidence bears examination.

The initial research paradigm employed to investigate own-race bias was the facial-recognition study. Black and White subjects are initially exposed to a set of critical slides of Black and White faces. After a short retention interval, subjects are tested for their ability to recognize the critical slides, using a signal-detection task. The critical slides are randomly distributed among a set of distracter slides. Each subject's task is to respond yes to a critical slide and no to a distracter slide. From the subject's responses to the set of critical and distracter slides, a recognition-ability score is calculated. This score is usually represented as d-prime, an index that gives the subject credit for responding yes to a critical slide but penalizes the subject for responding yes to a distracter slide. Own-race bias exists if d-prime (or recognition ability) is greater for own-race slides than it is for other-race slides.

An effect-size statistic can be used to examine the extent of own-race bias among Black and White subjects in a given study. Effect size represents the difference between the d-prime mean for own-race faces and the d-prime mean for other-race faces in standard deviation units. For example, Malpass and Kravitz (1969) presented Black and White psychology students at the University of Illinois with slides of 10 Black faces and 10 White faces. Later, subjects were asked to identify these slides out of a set of 80 slides that included the critical slides, 30 additional White distracters, and 30 additional Black distracters. They correctly identified 7.92 of the White subjects exhibited own-race bias. They correctly identified 7.92 of the critical White faces but falsely identified 3.46 of the distracter White faces, resulting in a d-prime of 2.14. They also correctly identified 6.08 of the critical Black faces but falsely identified 4.85 of the distracter Black faces, resulting in a d-prime of 1.18. Own-race bias is implicated by the larger d-prime obtained with own-race faces than with other-race faces. In this case, the effect size is 1.41, indicating that mean d-prime for own-race faces is 1.41 standard deviations greater than mean d-prime for other-race faces.

In order to examine the extent of own-race bias across the known set of facial-recognition studies, a meta-analysis was conducted. The analysis was conducted on those studies that exposed both White and Black subjects to slides of White and Black faces. Fourteen different samples from 11 different studies were employed in the analysis. The effect size of the own-race bias was calculated for Black and White subjects in each sample. These effect sizes were then averaged, so that the magnitude of own-race bias among Black and White subjects could be examined.

META-ANALYSIS

Data Search

Two reviews of the research on own-race bias in the facial-recognition literature were utilized (Lindsay & Wells, 1983; Malpass, 1982). The Lindsay

and Wells review involved 11 different studies: Brigham and Barkowitz (1978); Chance, Goldstein, and McBride (1975); Cross, Cross, and Daly (1971); Feinman and Entwisle (1976); Galper (1973); Goldstein and Chance (1976); Luce (1974); Malpass (1974); Malpass, Lavigueur, and Weldon (1973); and Shepherd, Deregowski, and Ellis (1974). We obtained copies of all these papers and calculated effect sizes for White and for Black subjects. Studies by Cross et al. (1971), Luce (1974), and Malpass (1974) were not included because not enough data were included in the original manuscript and the raw data were no longer available. The study by Goldstein and Chance (1976) was not included in the analysis because only White subjects were used.

Malpass's (1982) review included three studies not included in the Lindsay and Wells (1983) review: Brigham and Williamson (1979), Devine and Malpass (1981, 1985) and Ellis and Deregowski (1981). In addition, we included two studies reported in Barkowitz and Brigham (1982) that were not included in either the Lindsay and Wells or the Malpass review. Three studies (Barkowitz & Brigham, 1982; Ellis & Deregowski, 1981; Malpass & Kravitz, 1969) involved two separate samples. This resulted in the 14 samples listed in Table 1.

RESULTS

Effect sizes and sample sizes for Black and White subjects for each sample are presented in the upper portion of Table 1. These data were used to calculate the mean effect size (d), the variance of the effect sizes (S^2_d), the variance expected on the basis of sampling error (σ^2_e), and the standard deviation corrected for sampling error (σ_Δ), using the formulas given by Hunter, Schmidt, and Jackson (1982). These statistics are reported separately for Black and White subjects in the lower portion of Table 1. An effect size of $d = 0.00$ would indicate that faces of both races were recognized equally well. A positive effect size indicates that own-race faces were recognized better than other-race faces, whereas a negative effect size indicates that other-race faces were recognized better than own-race faces.

Both Black subjects ($d = .71$) and White subjects ($d = .69$) revealed an overall tendency to recognize own-race faces better than other-race faces. Converting these effect sizes to r, according to the formula:

$$r = d \div \sqrt{d^2 + 4},$$

we find that own-race bias accounts for 11% of the variance in recognition ability of Black subjects and 10% of the variance in recognition ability of White subjects.

Although the overall mean effect size for Black subjects is statistically equivalent to that of Whites $r(26) = .14$, $n.s.$, the confidence interval for Black

TABLE 1   Own Versus Other Effect Sizes for Black and White Subjects

Study	Sample Size		Effect Size	
	Black	White	Black	White
Barkowitz and Brigham, 1982 (Study 1)	81	174	-.19	.76
Barkowitz and Brigham, 1982 (Study 2)	58	43	.56	.45
Brigham and Barkowitz, 1978	86	76	1.247	.757
Brigham and Williamson, 1979	27	14	1.986	.347
Chance, Goldstein, and McBride, 1975	48	48	.596	.685
Devine and Malpass, 1985	24	24	1.127	1.084
Ellis and Deregowski, 1981 (transformed)	48	48	1.575	.741
Ellis and Deregowski, 1981 (untransformed)	48	48	.556	.185
Feinman and Entwisle, 1976	144	144	.436	.556
Galper, 1973	16	14	1.317	1.245
Malpass and Kravitz, 1969 (Illinois)	13	13	-.115	1.405
Malpass and Kravitz, 1969 (Howard)	7	7	-.107	1.343
Malpass, Lavigueur, and Weldon, 1973	61	67	.908	.364
Shepherd, Deregowski, and Ellis, 1974	32	32	.570	1.504

Note: $\bar{d} = .7087$ $.6848$; $S_d^2 = .3041$ $.0881$; $\sigma_e^2 = .0152$ $.0137$; $\sigma_\delta = .5375$ $.2728$.

subjects is wider than for Whites and does include zero: C.I. (95%) = (−.34 to 1.76) for Blacks and C.I. (95%) = (.16 to 1.22) for Whites. Although both groups are quite variable (even after correcting for sampling error), Black ds ($\sigma_\Delta^2 = .29$) are significantly more variable than White ds ($\sigma_\Delta^2 = .07$), $F(14, 14) = <.01$. This is due primarily to the data reported for Black subjects by Malpass and Kravitz (1969) and Barkowitz and Brigham (1982, Study 1), wherein Blacks performed better on White faces than Black faces. It should be noted that in both of Malpass and Kravitz's (1969) samples (students at Illinois and Howard Universities), a main effect for race of picture was significant: The White stimuli were more easily recognizable. However, the main effect for race of picture occurred only in the first Barkowitz and Brigham study, wherein both Blacks and Whites were superior at recognizing White faces. In their second study,

Barkowitz and Brigham found a crossover interaction and no main effect for race of picture. Both Blacks and Whites exhibited a strong own-race bias.

Across the 14 samples, the data from White subjects and Black subjects are quite variable, and in both cases the variation cannot be accounted for by sampling error alone. Sampling error accounts for only 16% of the variance among White ds and even less of the variance among Black ds. The remaining variation could be due to error of measurement, different standard deviations associated with recognition ability, or study artifacts such as differences among stimulus materials, length of the stimulus interval, length of the interstimulus interval, or retention interval. Different stimulus materials, in particular, may contribute to much of the variance in d-primes across studies. Distinctive faces may be easily recognized regardless of race. Samples of stimulus faces that include larger proportions of distinctive faces should yield lower cross-racial effect sizes. Malpass and Kravitz (1969) argued that this was the reason that they did not observe own-race bias in their Illinois and Howard University samples. There was a main effect for race of picture (both Blacks and Whites were superior at recognizing White faces), indicating perhaps that the White stimulus faces were more distinctive and easily recognizable than the Black stimulus faces.

DISCUSSION

This meta-analysis persuades us to disagree with those who have argued that the data from facial-recognition studies of own-race bias are equivocal and inconsistent. The data indicate that the own-race bias effect is quite consistent, in that it occurs for both Black and White subjects in 79% of the samples considered. The effect occurs with equal magnitude among both Black and White subjects, accounting for 11% of the variance in recognition ability of Black subjects and 10% of the variance in recognition ability of White subjects. Generalizing from facial-recognition studies to eyewitness identifications is problematic (Clifford, 1978). Eyewitness researchers recognize this, and they have begun conducting studies of the own-race bias employing more forensically relevant paradigms (Brigham, Maass, Snyder, & Spaulding, 1982; Platz & Hosch, in press). Both of these studies were conducted in convenience stores. An unusual interaction was staged between a customer who was working with the researchers and the store clerk. For example, the customer paid for a pack of cigarettes with all pennies and asked directions to the airport. Two hours later the clerk was interrogated by a "law intern" working with the researchers who showed the clerk a photographic lineup. Brigham and his colleagues found no support for own-race bias among either Black or White clerks in Tallahassee, Florida. One explanation for this finding might be the high rate of intergroup contact between convenience store clerks and their customers. Blacks represent 25% of the population in Tallahassee. On the other hand, Platz and Hosch found evidence of own-race bias among both Anglo-American and Mexican-American

convenience store clerks in El Paso, Texas, despite frequent intergroup contact. Mexican-Americans represent 63% of the population in El Paso. Some support for an intergroup-contact hypothesis was revealed by Platz and Hosch: both the Anglo-American and the Mexican-American clerks were particularly likely to misidentify a Black they had never seen before. Blacks represent only 3% of the population in El Paso.

Further research using more forensically relevant paradigms is necessary to improve the ecological validity of the literature on the own-race bias. Researchers conducting more forensically relevant research should be particularly attentive to target distinctiveness and attractiveness. Selection of a target who is either very distinctive or very attractive might artificially reduce the cross-racial effect. In addition, more research into the theoretical explanations for the effect is needed (Brigham & Malpass, 1985). The idea that own-race bias is a result of limited intergroup contact has received some support, but better measures of contact are needed.

REFERENCES

Barkowitz, P., & Brigham, J. C. (1982). Recognition of faces: Own-race bias, incentive, and time delay. *Journal of Applied Social Psychology, 12*, 255-268.

Brigham, J. C., & Barkowitz, P. (1978). Do "they all look alike?" The effect of race, sex, experience, and attitudes on the ability to recognize faces. *Journal of Applied Psychology, 8*, 306-318.

Brigham, J. C., Maass, A., Snyder, L. D., & Spaulding, K. (1982). Accuracy of eyewitness identifications in a field setting. *Journal of Personality and Social Psychology, 42*, 673-681.

Brigham, J. C., & Malpass, R. S. (1985). The role of experience and contact in the recognition of faces of own- and other-race persons. *Journal of Social Issues, 41*, 139-155.

Brigham, J. C., & Williamson, N. L. (1979). Cross-racial recognition and age: When you're over 60, do they still "all look alike?" *Personality and Social Psychology Bulletin, 5*, 218-222.

Chance, J., Goldstein, A. G., & McBride, L. (1975). Differential experience and recognition memory for faces. *Journal of Social Psychology, 97*, 243-253.

Clifford, B. R. (1978). A critique of eyewitness research. In M. Gruneberg, P. Morris, & R. Sykes (Eds.), *Practical aspects of memory* (pp. 199-209). London: Academic Press.

Cross, J. F., Cross, J., & Daly, J. (1971). Sex, race, age, and beauty as factors in recognition of faces. *Perception and Psychophysics, 10*, 393-396.

Devine, P. G., & Malpass, R. S. (1981). *Differential face recognition: A less than convincing explanation.* Paper presented at the meetings of the Psychonomic Society, Philadelphia.

Devine, P. G., & Malpass, R. S. (1985). Orienting strategies in differential face recognition. *Personality and Social Psychology Bulletin, 11*, 33-40.

Ellis, H. D., & Deregowski, J. B. (1981). Within-race and between-race recognition of transformed and untransformed faces. *American Journal of Psychology, 94*, 27-35.

Feingold, C. A. (1914). The influence of environment on identification of persons and things. *Journal of Criminal Law and Police Science, 5*, 39-51.

Feinman, S., & Entwisle, D. R. (1976). Children's ability to recognize other children's faces. *Child Development, 47*, 506-510.

Galper, R. E. (1973). Functional race membership and recognition of faces. *Perceptual and Motor Skills, 37*, 455-462.

Goldstein, A. G., & Chance, J. (1976). Measuring the psychological similarity of faces. *Bulletin of the Psychonomic Society, 7*, 407-408.

Hunter, J. E., Schmidt, F. L., & Jackson, G. B. (1982). *Meta-analysis: Cumulating research findings across studies.* Beverly Hills, CA: Sage.

Lindsay, R. C. L., & Wells, G. L. (1983). What do we really know about cross-race eyewitness identification? In S. M. A. Lloyd-Bostock & B. R. Clifford (Eds.), *Evaluating witness evidence: Recent psychological research and new perspectives* (pp. 219-234). New York: John Wiley.

Luce, T. S. (1974, November). Blacks, whites, and yellows: They all look alike to me. *Psychology Today*, pp. 105-108.

Malpass, R. S. (1974). Racial bias in eyewitness identification. *Personality and Social Psychology Bulletin, 1*, 42-44.

Malpass, R. S. (1982, March). *Differential recognition for faces of own and other race: A data summary.* Paper presented at the meetings of the American Academy of Criminal Justice Science, Louisville.

Malpass, R. S., & Kravitz, J. (1969). Recognition of faces of own and other race. *Journal of Personality and Social Psychology, 13*, 330-334.

Malpass, R. S., Lavigueur, H., & Weldon, D. E. (1973). Verbal and visual training in face recognition. *Perception and Psychophysics, 14*, 285-292.

Platz, S. J., & Hosch, H. M. (in press). Crossracial/ethnic identification: A field study. *Journal of Applied Social Psychology.*

Shepherd, J. W., Deregowski, J. B., & Ellis, H. D. (1974). A cross-cultural study of recognition memory for faces. *International Journal of Psychology, 9*, 205-211.

Yarmey, A. D., & Jones, H. P. T. (1983). Is eyewitness identification a matter of common sense? In S. M. A. Lloyd-Bostock & B. R. Clifford (Eds.), *Evaluating witness evidence: Recent psychological research and new perspectives* (pp. 13-40). New York: John Wiley.

Robert K. Bothwell teaches psychology at Pan American University in South Texas. His research focuses on factors associated with eyewitness accuracy. In addition to his work on cross-racial identification, he has been concerned with the effects of arousal on eyewitness memory and the relationship between eyewitness confidence and accuracy.

John C. Brigham is Professor of Psychology at Florida State University. His research interests include factors affecting the accuracy of eyewitness identification and the application of social psychological concepts to legal questions.

Roy S. Malpass is Professor of Behavioral Science at SUNY, Plattsburgh. His main interests are in cross-cultural psychology and social memory—particularly facial memory and eyewitness identification.

Improving the Reliability of Eyewitness Identification:
Putting Context Into Context

Brian L. Cutler and Steven D. Penrod
University of Wisconsin–Madison

Todd K. Martens
Harvard University

We examined the effects of context reinstatement procedures on eyewitness identification accuracy. Subjects were 290 undergraduates who viewed a videotaped reenactment of a liquor store robbery and, in a later session, attempted to identify the robber from a lineup parade. Two types of context reinstatement procedures were examined together with eight encoding, storage, and retrieval variables manipulated within the stimulus videotape and the lineup procedures. Disguise of the robber impaired identification accuracy ($p < .05$). There was a significant interaction between disguise and the context reinstatement interview ($p < .01$) such that the context reinstatement interview had a stronger impact on identification accuracy in the high-disguise condition. Lineup cues interacted with lineup composition ($p < .05$), retention interval ($p = .01$), and exposure to mug shots ($p = .05$; although in a manner contrary to our expectation). These interactions indicated that lineup context cues improved identification accuracy in the high-similarity, 2-week retention interval, and no mugshots conditions.

The unreliability of eyewitness identification has been amply documented (Brigham, Maass, Snyder, & Spaulding, 1982; Clifford & Bull, 1978; Loftus, 1979; Penrod, Loftus, & Winkler, 1982; Yarmey, 1979). Experiments on eyewitness identification are typically designed to identify circumstances under which eyewitnesses are particularly fallible. Such research is valuable, as information gained from these experiments may enable police, prosecutors, defense attorneys, and juries to make more informed assessments of eyewitness identifications (Cutler, Penrod, & Martens, 1987; Cutler, Penrod, & Stuve, in press; Loftus, 1983; although see McCloskey & Egeth, 1983, for a dissenting view). It is perhaps equally important to identify procedures that might improve the reliability of eyewitness identifications. Despite the clear advantages to the criminal justice system of establishing such procedures, there has, until recently, been little research in this domain of eyewitness identification.

One promising approach to the task of improving the reliability of eyewitness identification and recall involves procedures designed to reinstate the context surrounding an event. According to the network theory of memory (Bower, 1981; Collins & Loftus, 1975), environmental, emotional, and other contextual and stimulus-relevant information are encoded into

memory, together with the to-be-remembered stimulus, as a set of nodes that are connected to the to-be-remembered stimulus through associative path links. Contextual cues to recognition prime alternative pathways to activate the node representing the to-be-remembered stimulus (henceforth referred to as the *stimulus node*). Sometimes an individual's information search fails to prime the necessary paths leading to the stimulus node (i.e., the individual fails to recall the necessary information). Under such circumstances, alternative pathways may be primed through the use of contextual cues, making it more probable that the stimulus node will be activated and the information in question recalled.

Given the theoretical rationale governing the effects of context cues, our research adopts a rather liberal definition of context. We define context as any information that is encoded together with the to-be-recognized stimulus. The information is presumed to be stored in the memory network and connected to the to-be-recognized stimulus through associative pathways (cf. Tulving & Thomson, 1973).

Although in theory, the network model should hold for both recall and recognition, the effects of context reinstatement procedures are noted for being stronger in tests of recall than in tests of recognition (Bower, 1981; Smith, in press). Consistent with this conclusion, experiments that have tested the effect of context reinstatement on identification accuracy have yielded mixed results. Some investigators (e.g., Krafka & Penrod, 1985; Malpass & Devine, 1981a) have shown positive effects for context reinstatement, whereas others (e.g., Cutler, Penrod, O'Rourke, & Martens, 1986; Davies & Milne, 1985; Sanders, 1984) have shown null or weak results for context reinstatement. It is difficult to precisely determine the locus of the differential effects for context reinstatement. Comparisons across experiments are made difficult due to differences in retention interval, type of incident, and context reinstatement procedure, all of which could conceivably mediate the influence

of context reinstatement procedures. Note that the mixed results for context reinstatement are unlikely to be attributable to Type I error, as a recent meta-analysis of the facial recognition literature (Shapiro & Penrod, 1986) reveals that, across a large number of studies, context cues improve identification accuracy.

Why might encoding, storage, and retrieval factors mediate the influence of context cues on eyewitness identification accuracy? One plausible answer is Smith's (in press) *outshining hypothesis.* The outshining hypothesis acknowledges that in a recognition test certain aspects of the stimulus itself serve as context cues to recognition. In Smith's terms, if information provided by the stimulus serves as a strong cue to recognition, then any additional incidentally encoded context cues are "outshined." In terms of the network theory, the cues provided by the stimulus itself are often sufficient to activate the necessary paths to the stimulus node, so additional attempts to activate alternative pathways to the stimulus node through the use of context cues yield few gains in recognition accuracy (see also Bower, 1981). If, on the other hand, the stimulus properties that ordinarily cue recognition are degraded or absent and, consequently, the appropriate pathway to the stimulus node is not activated, then additional context cues improve recognition accuracy by activating alternative pathways to the stimulus node. In the case of eyewitness identification, facial characteristics and hair qualities are known to be important cues. If these cues are optimally encoded, then other contextually reinstated cues may yield little or no improvement in identification accuracy. But if these important cues are degraded due to impoverished encoding (e.g., disguise, weapon focus), storage factors (e.g., exposure to mug shots, long retention interval), or retrieval conditions (e.g., variations in lineup procedures, suggestive instructions), then contextually reinstated cues may enhance the accuracy of identifications by facilitating the priming of alternative paths to the individual's memorial representation of the to-be-identified target.

In summary, the mixed results for context reinstatement procedures might be due to differences in the qualitative and quantitative nature of context reinstatement procedures used in previous research or to encoding, storage, and retrieval factors that for theoretical reasons might moderate the effectiveness of context reinstatement procedures. An experiment was carried out to examine the influence of two types of context reinstatement procedures and the interactions between these procedures and a variety of encoding, storage, and retrieval factors on eyewitness identification accuracy. Because our primary concern is with techniques that can be used for actual eyewitness situations, we limit our study of contextual cues to those that could conceivably be implemented within current police investigatory procedures.

A second important concern of the present research is the predictive validity of eyewitness confidence. It is often concluded that the correlation between confidence and identification accuracy is weak (Bothwell, Deffenbacher, & Brigham, 1986; Wells & Murray, 1984). Wells and Lindsay (1985) raised criticisms about the manner in which confidence has been assessed in the eyewitness studies; for example, they point out that confidence is generally measured with a single item and might therefore suffer from unreliability (although cf. Cutler, Penrod,

O'Rourke, & Marten, 1986; Cutler, Penrod, & Martens, 1987; Murray & Wells, 1982). We attempt to improve on earlier assessments by using multiple confidence indices. In addition, we assess both confidence in the ability to correctly identify the perpetrator (before subjects view the lineup), as well as confidence in the lineup judgment.

Overview of the Experiment

Subjects in the present experiment viewed a videotaped robbery of a liquor store, and later attempted to identify the robber from a lineup parade. In light of Smith's (in press) outshining hypothesis, we manipulated variables that have been shown to affect identification accuracy. Our general expectation was that context reinstatement procedures will be effective in improving identification accuracy in circumstances under which identification accuracy is less reliable due to factors such as disguises worn by the robber, mugshot searches, the presence of a weapon, and substantial retention interval. Two types of context reinstatement procedures were examined; one procedure involved prelineup interviews and the other involved exposing subject–witnesses to context cues embedded within the lineup itself.

Context Reinstatement Interview

The context reinstatement interview was modeled after the procedures used by Krafka and Penrod (1985) and by Cutler, Penrod, O'Rourke, & Martens (1986, Experiment 2). Like Krafka and Penrod, we attempted to reinstate context by using a guided interview consisting of mnemonic procedures ("mnemonic instructions") developed by Geiselman, Fisher, MacKinnon, and Holland (1985), and by exposing subjects to a series of snapshots depicting the victim of the robbery, as well as the environment in which the robbery occurred (snapshot review). The mnemonic instructions are comprised of the following procedures: (a) mental reinstatement of the context surrounding an incident, (b) report of all information recalled, (c) rehearsal of events in different orders, and (d) rehearsal of events from different perceptual perspectives. These mnemonic instructions alone have been shown to enhance the accuracy of eyewitness recall (Geiselman et al., 1985). The snapshot display consisted of photographs of the inside of the liquor store and of the clerk behind the counter. A third set of contextual cues was added to this procedure. Subjects were instructed to reread their written descriptions of the robbery and of the physical characteristics of the robber that they completed immediately after viewing the videotaped robbery (as in Cutler, Penrod, O'Rourke, & Martens, 1986, Experiment 2); this procedure is referred to as *original description review.* The mnemonic instructions, the snapshot review, and the original description review were combined into a single interview procedure and are henceforth referred to as the *context interview.*

Lineup Context Cues

The second type of context reinstatement procedures examined in the present experiment involved physical characteristic context cues tested by Cutler, Penrod, O'Rourke, & Martens

(1986, Experiment 1). In their experiment, subjects were systematically exposed to cues such as voice, gait, posture, skin color, and a three-fourths pose, and these cues were manipulated separately. In the present experiment, four of these cues—voice, gait, posture, and three-fourths pose—were combined into a single variable referred to as *lineup context cues* (all subjects were shown color lineups). Thus, one half of the subjects were shown a lineup consisting of snapshots and slides of a front and full profile of the head and shoulders only. The other half were shown the same lineup features, but were also shown the three-fourths pose and the full view of the suspects' bodies (posture cues) from the three poses. Subjects in this condition were also shown a videotaped segment of each lineup suspect walking in and out of the room in which the lineup was held (gait cues), and heard voice samples from each suspect, which consisted of a single spoken line (voice cues). Viewing time of each lineup suspect was, of course, held constant. Although these lineup features are somewhat different from traditional context reinstatement manipulations, they fit the broad definition of context cues specified earlier. Other researchers have also manipulated pose as a means to reinstating context (e.g., Thomson, Robertson, & Vogt, 1982).

In addition to the two types of context reinstatement manipulations, eight other encoding, storage, and retrieval variables were manipulated. These variables were chosen because they have been shown to affect identification accuracy in previous research. A description of each follows.

Disguise. In one half of the videotaped robberies, the robber wore a hat fully covering his hair, and in the other half, the robber wore no hat.

Weapon visibility. In one half of the videotaped robberies, the robber outwardly brandished his handgun during the entire robbery, whereas in the remaining versions, the robber's handgun remained hidden throughout the robbery.

Retention interval. Subjects attempted an identification after either 2 days or 2 weeks.

Exposure to mugshots. One half of the subjects in the present experiment searched a series of 41 mug-shot slides (which included neither the robber nor the other lineup members) for the robber during the encoding session of the experiment, and the other half viewed no mug shots.

Lineup instructions. Before viewing the lineup parade, one half of the subjects were given instructions that explicitly offered the option of rejecting the lineup. The remaining subjects were given instructions that failed to explicitly offer this option. Failing to offer the option of rejecting the lineup typically increases false identification (Cutler, Penrod, & Martens, 1987; Malpass & Devine, 1981b).

Lineup type. One half of the subjects attempted an identification from an offender-present lineup, whereas the remaining half attempted an identification from an offender-absent lineup.

Lineup size. One half of the subjects viewed 6-suspect lineups, and the other half viewed 12-suspect lineups.

Lineup composition. In the present experiment, lineups were created on the basis of similarity ratings obtained in pilot work; these lineups were intended to differ with respect to fairness. Subjects viewed either high-similarity lineups (lineups that contained several members who resembled the robber in appear-

ance) or low-similarity lineups (lineups that contained few members who resembled the robber in appearance).

Method

Subjects

Subjects ($N = 290$) were volunteers from the University of Wisconsin–Madison introductory psychology subject pool, who received extra credit points for their participation. Subjects were randomly assigned to conditions. There were from 2 to 4 subjects per cell. Each cell necessitated a separate encoding and retrieval session, and where possible, subjects within the same cell were run in groups.

Design

In all, 10 variables (two levels each) were manipulated within a $2^{(7+3)}$ fractional factorial design. A fractional factorial design (Kenny, 1985) is one in which some main effects are deliberately confounded with higher order interactions between other main effects. In the present experiment, 7 variables were fully crossed with one another in a 128-cell design, whereas another 3 variables were confounded with higher order four- and five-way interactions. The advantage to this design is that 10 main effects, 45 two-way interactions, and three-way interactions of interest can be meaningfully assessed within a $2^7 = 128$-cell design. Testing 10 main effects and 45 two-way interactions in a full factorial design would necessitate $2^{10} = 1,024$ cells and many more subjects. The drawback to this fractional factorial design is that experimental efficiency is traded off against the fact that a few high-order interactions (four- and five-way) are confounded with main effects and there are many confounds among higher order interactions. Of course, these interactions tend to be uninterpretable in any event.

Materials

Stimulus videotapes. The plot of the vignette concerned a female clerk at a liquor store who, shortly after serving one customer, is confronted and robbed by a young man brandishing a handgun. The robber enters the store, demands the money from the register, and threatens to shoot the clerk. In the course of the interaction, the robber fires his weapon into the floor and roughs up the clerk. The entire videotape lasted approximately 100 s, and the robbery itself lasted approximately 75 s. Two variables, disguise and weapon visibility, were fully crossed within the videotapes. This was accomplished by a combination of repeated filming and editing of the robbery. Thus, in all, four videotapes were used. Stimulus videotapes were high quality, ¾-in. videocassettes, and were shown on a large (64-in. diagonal) projector screen, using a Kloss Nova Beam, Model 2.

Instructions. All of the instructions were given in writing. Subjects in the mug-shot condition were instructed to study each of the mug shots and to search for the robber, while the experimenter read a number aloud for each mug shot. Immediately following the presentation of the mug shots, subjects were further instructed to indicate the number of the mug shot that they thought was the robber or indicate that the robber was not among the mug shots (unbiased instructions).

Subjects in the context interview condition first received written mnemonic instructions to think back through the event, from beginning to end, and then in different orders and from different perspectives. Subjects were also instructed to try to remember the emotions they felt during the robbery and recall everything they viewed. Subjects were given up to 5 min to reminisce.

Lineup materials. All of the subjects were given a set of color snapshots of each lineup suspect. Subjects studied a front view (face and shoulders) of each suspect on one side of the snapshot and a full profile

B. CUTLER, S. PENROD, AND T. MARTENS

view on the reverse side of the snapshot. After subjects were given ample time (by their own acknowledgment) to examine the photo spread, subjects were shown a more detailed set of slides of each suspect. Only one suspect appeared on each slide. Subjects were allowed to study the photographs during the slide presentation in order to make comparisons among the suspects.

Subjects in the weak lineup context condition were shown slides of each suspect in front and full profile view, and slides consisted of views of the face and shoulders only. Subjects in the strong lineup context condition were shown slides of each suspect in front, three-fourth, and full profile views. In addition to views of the face and shoulders, full-body views of each suspect were also shown. Subjects in the strong lineup context condition also viewed a videotaped segment (¾ in. shown on a 25-in color monitor) of each lineup member walking in and out of the room in which the photographs and slides were taken, and heard a sample of each suspect's voice. The voice sample consisted of a single line spoken about the weather, and each suspect spoke the same line. The presentation of the enhanced-lineup condition proceeded as follows: Subjects were shown a series of slides of a given suspect immediately followed by the videotaped segment. Voice samples were given during the videotaped segment. Presentation time of each suspect was held constant (35 s) across lineups.

Interrogation questionnaire. The description of the robber and the robbery, which all subjects completed but only one half of the subjects reread before seeing the lineup, was given in response to a series of 40 probes about the event. In addition, subjects completed a checklist containing nine categories of body build (e.g., slender, stocky), nine hair colors, nine hair styles (e.g., shoulder length, balding), six eye colors, seven facial hair categories, and nine overall descriptions (e.g., well kept, slouched, ugly). No restrictions were placed on the number of categories that subjects were allowed to check.

Confidence questionnaires. Confidence in lineup choice was assessed twice. Immediately after viewing the stimulus videotape during the encoding session, subjects completed a *prejudgment confidence* questionnaire that consisted of the following two questions: "If we showed you a lineup in which the robber was present, how confident are you that you could choose the right person?" and "If we showed you a lineup in which the robber was not present, how confident are you that you would not mistakenly choose somebody out of the lineup?" Responses to both inquiries were indicated on 9-point scales ranging from *not at all confident* (1) to *very confident* (9). Immediately after rendering a judgment on the lineup task, subjects completed a *postjudgment confidence* questionnaire that consisted of the following three questions: "How confident are you that your choice is correct?" "How willing are you to sign a sworn statement that your choice is correct?" and "What is the probability that your choice is correct?" Responses to the first two questions were given on 9-point scales ranging from *not at all confident* and *not at all willing* (1) to *very confident* and *very willing,* respectively (9). Responses to the third inquiry were open-ended and could theoretically range from 0 to 1.00.

Procedure

During the encoding session, subjects were first shown one of the four videotaped robberies. Subjects completed the prejudgment confidence questionnaire, and then completed a questionnaire in which they were asked to describe the robbery and the physical characteristics of the robber. Subjects who were not in the mug-shot condition were excused at this point. Subjects in the mug-shot condition were given the mug-shot instructions and were then shown the mug shots. After completing the mug-shot procedure, subjects were excused.

At the beginning of the retrieval session, subjects in the context reinstatement condition were administered the reinstatement procedures in the following order: mnemonic instructions, original description review,

and snapshot display. Subjects who were not in the context reinstatement condition were given an innocuous imagery assessment questionnaire. Subjects were then given lineup instructions (biased or neutral) and were handed the photographs of the lineup suspects. After subjects indicated that they had enough time to study the photographs, subjects were shown the lineup. At the completion of the lineup phase, subjects indicated their judgments privately and completed the postjudgment confidence questionnaire.

Results

Overall, 234 subjects (81%) made a positive identification, whereas 56 subjects (19%) rejected the lineup. Of the 140 subjects who were shown offender-present lineups, 90 (64%) correctly identified the robber, 37 (27%) mistakenly identified a foil, and 13 (9%) incorrectly rejected the lineup. Of the 150 subjects shown an offender-absent lineup, 43 (29%) correctly rejected the lineup, whereas 107 (71%) falsely identified a foil. Collapsed across lineup type, the overall correct performance rate (CP; proportion of hits + proportion of correct rejections) was .46.

Lineup Decisions

Malpass and Devine (1984) argued that when examining the effects of variables on identification accuracy, it is informative to first account for the variable's effect on the lineup decision (i.e., the decision to identify a suspect or to reject the lineup), and through the consequence of that decision explain identification accuracy. This approach is especially appropriate if the effects of a variable on identification accuracy should theoretically be mediated through its effect on the lineup decision. Lineup instructions is the one factor in our experiment that meets this criterion. The remaining factors are hypothesized to affect identification accuracy by influencing sensitivity without necessarily affecting the lineup decision.

Thus, before analyzing the effects of lineup instructions on identification accuracy, we first analyzed its effects on the lineup decision. For the purpose of brevity, we limited our exploratory analyses of lineup decisions to main effects for each variable and the nine interactions between lineup instructions and each of the remaining factors. For this analysis, all positive identifications were scored 1 and lineup rejections were scored 0. A hierarchical regression analysis was then performed with each of the 10 factors entered on the first step and the 9 two-way interactions entered on the second step. The dependent variable

Table 1
Lineup Decisions: Predictors

Variable	M	d	t
Lineup type			
Offender absent	.72	.49	4.22*
Offender present	.90		
Lineup instructions			
Neutral	.69	.65	5.43*
Biased	.93		

Note. Means represent proportion of positive identifications.
* $p < .01$.

Table 2
Lineup Decisions: Summary Statistics

Step	R	Adjusted R^2	F total	df	MS_e	F change	df
1	.400	.130	5.33*	10, 279	.136		
2	.431	.129	3.25*	19, 270	.136	.95	9, 270

* $p < .01$.

was lineup decision (positive identification or lineup rejection). A summary of this regression analysis appears in Tables 1 and 2.

As expected, biased lineup instructions significantly increased the number of positive identifications. Of course, lineup type (offender present versus offender absent) also had a significant influence on the type of decision made. None of the other main effects or any of the two-way interactions were statistically significant.

Identification Performance

In order to examine the effects of the independent variables on identification performance, correct judgments were scored 1 and incorrect judgments were scored 0 (to form the CP score). With CP as the dependent variable, a hierarchical regression analysis was performed with the 10 predictor variables entered on the first step and the subsequent 45 two-way interactions entered on the second step. The 36 three-way interactions between lineup type (offender present vs. offender absent) and all other predictors were entered on the third step, to determine whether correct identifications and correct rejections were affected similarly by context and other variables. A summary of the regression results is displayed in Tables 3 and 4.

Significant main effects were found for lineup type and disguise of the robber. Fewer correct judgments were obtained in the offender-absent (as compared with offender-present) lineups and in the disguise (as compared with no-disguise) condition. Of the 45 two-way interactions, 6 were significant at $p < .05$, which is more than the number of significant interactions expected by chance alone.

Lineup context interacted significantly with lineup composition, such that context cues significantly improved performance if subjects were shown high-similarity lineups, but not if subjects were shown low-similarity lineups. Note that the difference in identification accuracy between the strong lineup cues–low similarity and strong lineup cues–high similarity cells (.54 vs. .44, respectively) is likely to be due to chance variation ($p > .05$).

Context cues in the lineup parade also interacted significantly with retention interval such that providing subjects with strong context cues in the lineup improved performance in the 2-week retention interval condition but had little effect on performance among subjects in the 2-day retrieval condition. Although it is apparent from the cell means that the strong cues in the lineup improved identification performance to a level above that obtained in the 2-day retention interval condition (.59 vs. .49), this difference is probably due to chance variation ($p > .05$).

The interaction between lineup context cues and exposure to mug shots was also significant. Contrary to our expectation, lineup context cues significantly improved identification accuracy among subjects who were not shown mug shots, but had little effect on identification performance among subjects who had been shown mug shots. The outshining hypothesis would lead us to expect the effects of context reinstatement to be stronger if subjects had been shown mug shots.

The context interview interacted significantly with disguise. It significantly improved identification performance if the robber had been disguised during the robbery, but had little effect on identification performance if the robber had not been disguised.

Disguise interacted significantly with lineup size such that disguise had a significantly larger effect on performance among subjects in the 12-suspect lineup condition than among subjects in the 6-suspect lineup condition.

Given that subjects in the biased lineup instruction condition have a strong tendency to make positive identifications (controlling for the presence of the offender in the lineup), biased lineup instructions should strongly reduce identification performance in the offender-absent condition, because any identification is an incorrect one. The effects of biased lineup instructions on identification performance in the offender-present condition should be less strong, because even though subjects are more inclined to make a positive identification, sometimes the identification is correct.

Biased instructions had a significantly larger impairment on identification performance in the offender-absent condition than in the offender-present conditions. This interaction might best be understood in terms of diagnosticity (Malpass & Devine, 1984; Wells & Lindsay, 1980). Diagnosticity refers to the ratio of correct identifications in offender-present lineups to false identifications in offender-absent lineups; therefore, higher ratios indicate better diagnosticity. Biased lineup instructions yielded a diagnosticity ratio of $.61/(1 - .13) = .70$, whereas neutral instructions yielded a diagnosticity ratio of $.67/(1 - .43) = 1.18$. Clearly, biased instructions strongly reduce the diagnosticity of identifications.

It is perplexing that the mug-shot manipulation had little effect on identification accuracy. One plausible hypothesis[1] for the lack of an effect is that the mug-shot search (in which the robber was *not* present) served to reinforce the subjects' beliefs that the robber might be absent from any additional set of mug shots or lineups that are to be searched. Such a hypothesis

---

[1] We are grateful to the anonymous reviewer for pointing out this plausible hypothesis.

Table 3
Identification Accuracy: Predictors

Variable	M	d	t
Step 1			
Lineup type			
Offender absent	.29	.77	6.52**
Offender present	.64		
Disguise			
Low	.51	−.24	−2.19*
High	.40		
Step 2			
Lineup Context × Lineup Composition			2.35*
Low Similarity			
Weak cues	.36	.40	3.06**
Strong cues	.54		
High Similarity			
Weak cues	.50	−.13	−0.99
Strong cues	.44		
Lineup Context × Retention Interval			2.91**
2-week retention interval			
Weak cues	.37	.49	3.75**
Strong cues	.59		
2-day retention interval			
Weak cues	.49	−.22	−1.68
Strong cues	.39		
Lineup Context × Exposure to Mug shots			−1.97*
No mug shots			
Weak cues	.35	.35	2.68**
Strong cues	.51		
Mug-shot search			
Weak cues	.51	−.09	−0.69
Strong cues	.47		
Context Interview × Disguise			2.87**
No disguise			
No interview	.57	−.22	−1.68
Interview	.47		
Disguise			
No interview	.29	.49	4.74**
Interview	.51		
Disguise × Lineup Size			−2.46**
12-suspect			
Low disguise	.51	−.31	−2.37**
High disguise	.37		
6-suspect			
Low disguise	.49	−.04	−0.31
High disguise	.47		
Lineup Instruction × Lineup Type			3.44**
Offender absent			
Neutral	.43	−.66	−5.04**
Biased	.13		
Offender present			
Neutral	.67	−.13	−0.99
Biased	.61		

Note. Means represent proportion correct.
* $p < .05$. ** $p < .01$.

would be supported by an interaction between lineup type (offender present vs. offender absent) and exposure to mug shots. The interaction, though, was nonsignificant.

The third step of the regression equation examined whether correct identifications and correct rejections were affected similarly by context and other independent variables. This was accomplished by testing the three-way interactions between lineup (offender present vs. offender absent) and each of two-way interactions. The number of significant three-way interactions (3) did not exceed chance levels; these interactions are therefore not discussed. None of the significant two-way interactions were further qualified by three-way interactions with lineup type. Note, however, that 6 of 36 three-way interactions were not entered into the equation because of minimal tolerance—they were too highly correlated with other interaction terms. Given that the interactions with which these terms are confounded were nonsignificant, there is little threat to the interpretation of our results.

Confidence

Two subjects failed to complete the prejudgment confidence questionnaire, 4 subjects failed to complete two of the postjudgment confidence questions, and 14 subjects failed to complete one postjudgment confidence question; their data were therefore excluded from the following analyses. Mean prejudgment confidence regarding the ability to correctly identify the robber was 7.30 ($SD = 1.35$), and mean prejudgment confidence regarding the ability to correctly reject the offender-absent lineup was 6.02 ($SD = 1.73$). The average confidence rating was 5.76 ($SD = 2.00$) for postjudgment confidence in judgment on the lineup task, 4.06 ($SD = 2.41$) for willingness to sign a sworn statement, and .56 ($SD = .28$) for probability of a correct judgment. Correlations between confidence measures, identification performance, and choosing, are displayed in Table 5.

The two measures of prejudgment confidence correlated significantly, $r = .37$, $p < .01$, and the correlations between each of these measures and performance (CP) were of similar magnitude. Prejudgment confidence regarding the ability to correctly identify the suspect correlated .10 with performance, whereas prejudgment confidence regarding the ability to correctly reject the lineup correlated .07 with performance. The three postjudgment confidence ratings correlated highly with one another—the average correlation was .72, $p < .01$—but minimally with prejudgment confidence measures—the average intercorrelation was .16, $p = .01$. Furthermore, the postjudgment confidence ratings correlated more highly with performance than did prejudgment confidence. For confidence in judgment, $r = .30$, $p < .01$, for willingness to sign a sworn statement, $r = .32$, $p < .01$, and for probability of correct judgment, $r = .27$, $p < .01$.

Clearly, pre- and postjudgment confidence are conceptually different measures; they demonstrate weak relations with one another, and they demonstrate different patterns of relations with identification performance. Given the internal consistency within these sets of measures, we decided to standardize and then aggregate prejudgment confidence measures to form a more reliable prejudgment confidence score and to standardize and then aggregate postjudgment confidence measures to form a more reliable postjudgment confidence score. The aggregate prejudgment confidence score (henceforth referred to as prejudgment confidence) correlated .10, $p = .05$, with performance, whereas the aggregate postjudgment confidence score

Table 4
Identification Accuracy: Summary Statistics

Step	R	Adjusted R^2	F total	df	MS_e	F change	df
1	.404	.134	5.47*	10, 279	.216		
2	.580	.180	2.15*	55, 234	.204	1.35	45, 234
3	.656	.193	1.81*	85, 204	.201	1.12	30, 204

* $p < .01$.

(henceforth referred to as postjudgment confidence) correlated .33, $p < .01$, with performance. The correlation between pre- and postjudgment confidence was .22, $p < .01$.

The estimator and system variables were also tested for their effects on confidence and as moderators of the confidence–accuracy correlation. The first analysis examined the aggregate prejudgment confidence as the dependent measure. Identification accuracy was entered on the first step, disguise and weapon visibility on the second, and the two interaction terms (Disguise × Identification Accuracy and Weapon Visibility × Identification Accuracy) were entered on the third step. The interaction terms test heterogeneity of variance, or the moderator effects. Disguise was a significant predictor of prejudgment confidence (semipartial $r = -.17$, $p < .01$). No other predictors were significant.

A similar analysis was then carried out with the aggregate postjudgment confidence as the dependent variable. Identification accuracy was entered on the first step, the 10 predictors on the second, and the 10 interactions between identification accuracy and each predictor were entered on the third step. Exposure to mug shots significantly reduced postjudgment confidence (semipartial $r = -.17$, $p < .01$). Subjects shown an offender-present lineup were significantly more confident than were subjects shown an offender-absent lineup (semipartial $r = .14$, $p < .05$). None of the moderator effects were statistically significant.

Discussion

Results of this experiment indicate that both context reinstatement interviews and lineup context cues affect identification performance by interacting with other variables that affect identification performance. The procedures differ, however, in the variables with which they interact. The effectiveness of the context interview was mediated by the disguise of the robber. This disguise, which consists of a hat covering most of the robber's hair, reliably reduces identification accuracy (Cutler, Penrod, O'Rourke, & Martens, 1986; Cutler, Penrod, & Martens, 1987; Shapiro & Penrod, 1986). The context reinstatement interview was more effective in the disguise condition (poor encoding conditions) than in the no-disguise condition.

The effects of lineup contextual cues were moderated by the similarity of the lineup members to the target, retention interval, and exposure to mug shots. Lineup context cues improved identification performance if subjects were presented with high-similarity lineups; that is, if lineup suspects resembled the offender in physical appearance. It is generally agreed that lineups should be fair tests (Doob & Kirshenbaum, 1973; Malpass & Devine, 1983; Wells, Leippe, & Ostrom, 1979) and should therefore contain foils who look like the suspect. Results of this experiment show that when lineups do contain such foils, physical characteristic context cues contribute to improved identification performance.

Retention interval is a strong predictor of recognition accuracy in facial recognition and eyewitness identification experiments (Shapiro & Penrod, 1986). As shown in the present experiment, the effectiveness of context reinstatement varied as a function of retention interval, thus complicating comparisons of context effects across experiments. This finding is especially noteworthy given that in the eyewitness identification literature, retention intervals vary from less than 1 hour (e.g., Sanders, 1984) to 5 months (Malpass & Devine, 1981a).

Table 5
Correlations Between Confidence and Identification Accuracy

Measure	1	2	3	4	5	6	7	8	9
Prejudgment confidence									
1. Correct identification	—	.37	.83	.24	.23	.16	.23	.10	-.01
2. Correct rejections		—	.82	.14	.11	.11	.13	.07	.03
3. Aggregate			—	.23	.21	.16	.22	.10	.01
Postjudgment confidence									
4. Decision specific				—	.80	.70	.92	.30	-.04
5. Willingness to sign a sworn statement					—	.66	.91	.27	-.07
6. Probability						—	.87	.27	.01
7. Aggregate							—	.33	-.04
Identification performance									
8. CP								—	-.30
9. Choosing									—

Note. $N = 274$. Correlations above .10 are significant at $p < .05$; correlations above .14 are significant at $p < .01$. CP = identification accuracy.

Lineup context cues improved identification performance among subjects who were not shown mug shots, but had less of an effect among subjects who were shown mug shots. This finding is perplexing, as it is opposite to the hypothesis derived from the outshining hypothesis.

With respect to eyewitness confidence, our results are consistent with those obtained in previous experiments (Cutler, Penrod, O'Rourke, & Martens, 1986; Cutler, Penrod, & Martens, 1987). In general, the confidence–performance relation was found to be significant, but moderate in size, and confidence in lineup decision was a much better predictor of identification performance than was confidence in ability to identify the offender. The finding that prejudgment confidence correlated so weakly with identification accuracy suggests that a witness's initial confidence in his or her ability to identify a perpetrator should not be used to support the validity of an identification. In addition, the finding implies that witnesses whose confidence in their ability to correctly identify a perpetrator wanes, might nevertheless be encouraged to attempt an identification.

Although the amount of variance in performance accounted for by confidence is not great, it is important to note that (a) as illustrated elsewhere (Cutler & Penrod, 1986; Nunnally, 1978), the point-biserial correlation is highly susceptible to attenuation due to differentially skewed distributions among the dichotomous variable (identification accuracy) and the continuous variable (confidence), and (b) the absolute value of a correlation coefficient may be a better estimate of strength of association than is the squared correlation coefficient (Ozer, 1985; see also Wells & Lindsay, 1985).

Together with the results of the Krafka and Penrod (1985) and Malpass and Devine (1981a) studies, the results of the present experiment should help to dispel some pessimistic notions regarding the effects of contextual cues on identification accuracy. We have found that physical characteristic cues in the lineup and reinstatement of environmental and emotional state context are effective in enhancing identification performance. However, our results, together with the other experiments reviewed, suggest that there must be an appreciable retention interval or impairment of memory due to other encoding, storage, or retrieval factors for context reinstatement to be effective. Physical characteristic context cues appear to be especially sensitive to retention interval. From a theoretical perspective, it would be helpful to know why some context cues are more effective than other context cues and why various context cues are differentially mediated by encoding, storage, and retrieval factors. From a forensic perspective, more research is needed to identify procedures that increase the reliability of eyewitness memory—in conjunction, naturally, with research that identifies conditions under which eyewitness memory is likely to be fallible. Reinstatement of context is a promising approach to the enhancement of eyewitness performance—one that surely deserves further attention in both applied and theoretical realms.

References

Bothwell, R. K., Deffenbacher, K. A., & Brigham, J. C. (1986). *Correlation of eyewitness accuracy and confidence: The optimality hypothesis revisited.* Manuscript submitted for publication.

Bower, G. H. (1981). Mood and memory. *American Psychologist, 36,* 129–148.

Brigham, J. C., Maass, A., Snyder, L. D., & Spaulding K. (1982). Accuracy of eyewitness identifications in a field setting. *Journal of Personality and Social Psychology, 42,* 673–680.

Clifford, B. R., & Bull, R. (1978). *The psychology of person identification.* London: Routledge & Kegan Paul.

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review, 5,* 85–88.

Cutler, B. L., & Penrod, S. D. (1986, May). *Eyewitness calibration: A within-subject perspective on the eyewitness confidence–accuracy relation.* Paper presented at the meeting of the Midwestern Psychological Association, Chicago.

Cutler, B. L., Penrod, S. D., & Martens, T. K. (1987). The reliability of eyewitness identifications: The role of system and estimator variables. *Law and Human Behavior, 11,* 233–258.

Cutler, B. L., Penrod, S., O'Rourke, T. E., & Martens, T. K. (1986). Unconfounding the effects of contextual cues on eyewitness identification accuracy. *Social Behavior, 1,* 113–134.

Cutler, B. L., Penrod, S. D., & Stuve, T. E. (in press). Juror decisionmaking in eyewitness identification cases. *Law and Human Behavior.*

Davies, G. M., & Milne, A. (1985). Eyewitness composite production as a function of mental or physical reinstatement of context. *Criminal Justice and Behavior, 12,* 209–220.

Doob, A. N., & Kirshenbaum, H. M. (1973). Bias in police lineups: Partial remembering. *Journal of Police Science and Administration, 1,* 287–293.

Geiselman, R. E., Fisher, R. P., MacKinnon, D. P., & Holland, H. L. (1985). Eyewitness memory enhancement in the police interview: Cognitive retrieval mnemonics versus hypnosis. *Journal of Applied Psychology, 70,* 401–412.

Kenny, D. (1985). Quantitative methods for social psychology. In G. Lindzey & E. Aronson (Eds.), *The handbook for social psychology* (Vol. 1, pp. 487–508). Hillsdale, NJ: Erlbaum.

Krafka, C., & Penrod, S. (1985). Reinstatement of context in a field experiment on eyewitness identification. *Journal of Personality and Social Psychology, 49,* 58–69.

Loftus, E. F. (1979). *Eyewitness testimony.* Cambridge, MA: Harvard University Press.

Loftus, E. F. (1983). Silence is not golden. *American Psychologist, 38,* 564–572.

Malpass, R. S., & Devine, P. G. (1981a). Guided memory in eyewitness identification. *Journal of Applied Psychology, 66,* 343–350.

Malpass, R. S., & Devine, P. G. (1981b). Eyewitness identification: Lineup instructions and the absence of the offender. *Journal of Applied Psychology, 66,* 482–489.

Malpass, R. S., & Devine, P. G. (1983). Measuring the fairness of eyewitness identification lineups. In S. M. A. Lloyd-Bostock & B. R. Clifford (Eds.), *Evaluating witness evidence* (pp. 81–102). New York: Wiley.

Malpass, R. S., & Devine, P. G. (1984). Research on suggestion in lineups and photospreads. In G. L. Wells & E. F. Loftus (Eds.), *Eyewitness testimony: Psychological perspectives* (pp. 64–91). New York: Cambridge University Press.

McCloskey, M., & Egeth, H. (1983). Eyewitness identification: What can a psychologist tell a jury? *American Psychologist, 38,* 550–563.

Murray, D. M., & Wells, G. L. (1982). Does knowledge that a crime was staged affect eyewitness accuracy? *Journal of Applied Social Psychology, 12,* 42–53.

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.

Ozer, D. J. (1985). Correlation and the coefficient of determination. *Psychological Bulletin, 97,* 307–315.

Penrod, S., Loftus, E. F., & Winkler, J. D. (1982). Eyewitness reliability.

EYEWITNESS IDENTIFICATION

In R. Bray & N. Kerr (Eds.), *The psychology of the courtroom* (pp. 119–168). New York: Academic Press.

Sanders, G. L. (1984). The effects of context cues on eyewitness identification responses. *Journal of Applied Social Psychology, 14,* 386–397.

Shapiro, P., & Penrod, S. D. (1986). A meta-analysis of the facial recognition studies. *Psychological Bulletin, 100,* 139–156.

Smith. (in press). Environment context-dependent memory. In G. Davies & D. Thomson (Eds.), *Memory in context: Context in memory.* Chichester, England: Wiley.

Thomson, D. M., Robertson, S. L., & Vogt, R. (1982). Person recognition: The effect of context. *Human Learning, 1,* 137–154.

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review, 80,* 352–373.

Wells, G. L., Leippe, M. R., & Ostrom, T. M. (1979). Guidelines for empirically assessing the fairness of a lineup. *Law and Human Behavior, 3,* 285–293.

Wells, G. L., & Lindsay, R. C. L. (1980). On estimating the diagnosticity of eyewitness nonidentifications. *Psychological Bulletin, 88,* 776–786.

Wells, G. L., & Lindsay, R. C. L. (1985). Methodological notes on the confidence–accuracy relationship in eyewitness identification. *Journal of Applied Psychology; 70,* 413–419.

Wells, G. L., & Murray, D. M. (1984). Eyewitness confidence. In G. L. Wells & E. F. Loftus (Eds.), *Eyewitness testimony: Psychological perspectives* (pp. 155–170). New York: Cambridge University Press.

Yarmey, A. D. (1979). *The psychology of eyewitness testimony.* New York: Free Press.

c

*338 EYEWITNESS IDENTIFICATION EVIDENCE

Evaluating Commonsense Evaluations

Jennifer L. Devenport [FNa1]

Steven D. Penrod

Brian L. Cutler

Although **eyewitness identifications** are among the most common forms of evidence presented in criminal trials, both archival studies and psychological research suggest that eyewitnesses are frequently mistaken in their identifications (B. L. Cutler & S. D. Penrod, 1995). In recognition of this problem, the legal system has established a number of safeguards to protect defendants from erroneous convictions resulting from mistaken identifications. These safeguards are based on assumptions regarding attorney, judge, and juror commonsense knowledge of the factors influencing **eyewitness identification** accuracy. This article addresses the validity of these assumptions by examining the role of commonsense knowledge in attorney, judge, and juror evaluations of **eyewitness identification** evidence. It concludes that, although these safeguards may not be as effective as the legal system intended them to be, there are a number of practices and policies that may be implemented to safeguard defendants further.

There is little question that **eyewitness identifications** are among the most important forms of evidence presented in criminal trials. The importance of these identifications is underscored by the fact that mistaken identifications appear to be the most frequent source of erroneous convictions. In their study of 28 cases of mistaken convictions in which defendants were subsequently cleared with scientific DNA evidence, Connors, Lundregan, Miller, and McEwen (1996) reported that all the convictions were predicated on mistaken **eyewitness identifications**. Furthermore, Huff (1987; see also Huff, Rattner, & Sagarin, 1996) implicated mistaken **eyewitness identifications** in 60% of the more than 500 erroneous convictions he studied.

One of the other ways by which it is known that eyewitness performance is a matter of serious concern in criminal cases is from the results of eyewitness studies conducted under fairly realistic conditions (e.g., Brigham, Maass, Snyder, & Spaulding, 1982; Krafka & Penrod, 1985; Pigott, Brigham, & Bothwell, 1990; Platz & Hosch, 1988). Although the central purpose of such studies has been to estimate the effects of particular factors on identification accuracy (witness and perpetrator race in Brigham et al. and Platz & Hosch; identification procedures in Krafka & Penrod; and the relation between accuracy of eyewitnesses' descriptions and their identifications in Pigott et al.), such studies are conducted in fairly realistic settings, and the identification accuracy rates they yield are one indication of the base rates likely to be found in cases involving actual **eyewitness identifications**.

*339 Brigham et al. (1982), for example, sent students into convenience stores and had them engage in transac-

tions lasting several minutes with store clerks. The clerks were later asked to identify the customers from photoarrays. A total of 73 clerks attempted identifications of the student customers after a 2-hr time delay; 50 of their 146 identifications from target-present arrays (34%) were correct. Krafka and Penrod (1985) used a similar procedure; their 85 clerks were shown either a customer-present or a customer-absent photoarray. When the customer was present in the photoarray, 41% of the clerks correctly identified him; whereas 34% falsely identified someone else's photograph as that of the customer when the customer was absent from the photoarray. Platz and Hosch (1988) also used a convenience store scenario, with three customers. In this study, each of 86 clerks attempted to identify all three customers. Overall, 44% of the identifications were correct. Pigott et al. (1990) used local banks for their study. Their customers interacted with tellers for about 90 sec-under rather suspicious circumstances. About 4 to 5 hr later the 47 tellers were shown a customer-present or customer-absent photoarray. Of the tellers shown a customer-present photoarray, 48% made a correct identification, whereas 38% made a false identification when shown a customer-absent photoarray.

In these four fairly realistic field experiments, data were gathered from 291 eyewitnesses who were given 536 separate identification tests. The average percentage of correct identifications was 42% when the customers were in the arrays, and the average percentage of false identifications was 36% when the customers were not in the arrays. In short, attempted identifications of individuals seen briefly in nonstressful conditions and after only short delays were frequently inaccurate: Witnesses often failed to identify targets when they were present and frequently identified innocent persons when targets were not present. Of course, **eyewitness identifications** are made under all sorts of witnessing and identification conditions, but these studies were based on conditions that are commonly encountered by eyewitnesses. When witnessing and identification conditions are better or worse than those found in these studies, it is likely that eyewitness performance improves or deteriorates.

Legal Safeguards Against Mistaken Convictions

Both the psychological and legal communities have recognized that there are problems with **eyewitness identification** evidence. Within the psychological community, substantial research efforts have been devoted to the task of discovering the sources of these errors and devising ways to minimize them (see Cutler & Penrod, 1995, for a review). Within the legal community, efforts have been made to protect defendants from erroneous convictions resulting from mistaken identifications. As part of these efforts, courts have designed several procedural safeguards intended to reduce mistaken identifications and convictions. These safeguards include such legal devices as the presence of counsel at postindictment, live lineups (*Kirby v. Illinois*, 1972; *United States v. Ash*, 1973; *United States v. Wade*, 1967), opportunities for motions to suppress identifications, cross-examination of identifying witnesses (Walters, 1985), and expert psychological testimony about factors that influence eyewitness memory (*People v. McDonald*, 1984). These safeguards are based on assumptions regarding attorneys', judges', and jurors' commonsense knowledge about factors that influence **eyewitness identifications**.

*340 As Finkel (1995) has underscored, there can sometimes be significant differences between the layperson's understanding and application of the law and trial evidence and those of the legal professional. Although Finkel's analysis emphasizes discrepancies between lay and professional understandings of the law and the influence of these discrepancies on juror's efforts to arrive at just decisions, our emphasis is on a somewhat different aspect of lay and professional common sense. This article addresses the effectiveness of safeguards against mistaken **eyewitness identifications** by examining the role of lay and professional commonsense (and legal) knowledge in the evaluation of **eyewitness identification** evidence. Although our focus is on juror commonsense knowledge, the application of juror common sense occurs in a context in which the legal and commonsense knowledge of attorneys and judges shapes and gives substance to the evidence evaluated by jurors. Fortunately, some empirical work exists that has examined attorney and judge commonsense knowledge of the factors that influence **eyewitness identification** performance, and we consider that research in our review and compare and contrast it with juror commonsense knowledge. We then examine the impact of scientific knowledge, in the form of expert psychological testimony, on jurors' understanding of and evaluation of **eyewitness identification** evidence.

Of particular interest in this enterprise are the divergences among common, legal, and scientific knowledge

concerning **eyewitness identifications**. Assuming that systematic scientific knowledge of the factors that influence **eyewitness identifications** is the gold standard of knowledge, how do attorneys', judges', and jurors' knowledge fare in comparison? As is shown below, although the answers are worrisome, the situation is far from hopeless; sound policies and practices offer the prospect of closing the gap between commonsense and scientific knowledge.

Attorneys' Commonsense Knowledge About Factors That Influence Eyewitness Performance

The presence of counsel safeguard serves to protect defendants from mistaken identifications by allowing defense attorneys to be present at postindictment live lineups. Attorneys who are present at postindictment lineups may advise their clients of their rights and obligations, oppose the use of suggestive identification procedures, and record suggestive identification procedures for the purpose of later filing a motion to suppress the identification (Stinson, Devenport, Cutler, & Kravitz, 1996). For the presence of counsel safeguard to be effective, however, attorneys must be aware of the factors that influence **eyewitness identification** accuracy in general and lineup suggestiveness in particular.

Surveying Attorney Knowledge

Few studies have assessed attorney commonsense knowledge regarding factors influencing **eyewitness identification** accuracy (Brigham & Wolfskiel, 1983; Rahaim & Brodsky, 1982; Stinson et al., 1996), and of these studies only one has assessed whether attorneys possess commonsense knowledge regarding factors that influence lineup suggestiveness (Stinson et al., 1996). Rahaim and Brodsky (1982) surveyed 42 attorneys and assessed their knowledge of factors influencing **eyewitness identification** accuracy. Overall, the results revealed that attorneys were generally sensitive to the effects of race and of stress or violence on **eyewitness identification** accuracy but appeared to believe mistakenly that there is *341 a strong relation between eyewitness confidence and identification accuracy (Penrod & Cutler, 1995).

In a larger scale study, Brigham and Wolfskiel (1983) surveyed 235 attorneys and assessed their knowledge regarding the factors they believed most likely to affect **eyewitness identification** accuracy. These attorneys reported being involved with eyewitnesses with some regularity. Photoarrays were encountered at least once a week by 59% of prosecutors and 25% of defense attorneys, and live lineups were encountered at least once a week by 23% of prosecutors and 9% of defense attorneys. Significantly more prosecutors (84%) than defense attorneys (36%) believed that "90% or more" of identifications are probably correct. Prosecutors (75%) were also more likely than defense attorneys (56%) to believe that eyewitnesses more commonly fail to identify guilty suspects than falsely identify innocent ones. The survey also contained questions about factors that may be perceived to influence identification accuracy. As is consistent with the empirical literature, both groups of attorneys believed cross-race identifications to be less accurate than same-race identifications even though their estimates of the overall accuracy rates differed markedly. Furthermore, more than 60% of attorneys thought that males and females would perform comparably on identification tests and that more intelligent witnesses were more likely to be accurate than less intelligent witnesses. Most attorneys, however, thought that education was not related to identification accuracy. These beliefs are consistent with the psychological literature with respect to gender and education (Cutler & Penrod, 1995), but there is little evidence to support a relation between intelligence and identification accuracy (Cutler & Penrod, 1989).

The attorneys were also asked, in open-ended format, what factors they believed to be related to identification accuracy. The characteristics most frequently mentioned by defense attorneys were physical characteristics of the suspect (60%), lighting at the scene of the crime (39%), exposure duration during the crime (36%), proximity to the suspect during the crime (34%), and physical appearance of the suspect's body (33%). The characteristics most commonly mentioned by prosecutors were physical characteristics of the suspect (68%), lighting at the scene of the crime (60%), exposure duration during the crime (52%), whether the witness had a good memory (32%), and physical appearance of the suspect's body (32%). Most of these factors are probably important.

Overall, the factors mentioned by attorneys tended to be predictive of identification accuracy, but a number of other potentially important factors were not mentioned: Weapon focus, disguises, changes in facial features, cross-race recognition, retention interval, and factors associated with the suggestiveness of identification tests did not rank on either group of attorneys' "top 10" lists. It is possible, however, that these factors present themselves with sufficient rarity that attorneys do not think of them or think they are less important than those factors they did mention. It is notable that there were substantial differences between defense and prosecuting attorneys concerning their top 10 factors. Across the two lists, only the physical characteristics of the suspect and lighting at the scene of the crime were mentioned by a majority of the respondents.

Defense and prosecution attorneys also had different perceptions of how eyewitness testimony is used in court. When asked about how much weight judges and juries give to eyewitness evidence, 89% of defense attorneys but only 7% of *342 prosecutors indicated that it was "too much." When asked whether "a psychologist's expert opinion should be considered in court when deciding the reliability of **eyewitness identification**," 11% of defense attorneys replied "never," 30% replied "rarely or only in unusual cases," 32% replied "fairly often," and 27% replied "routinely." For prosecutors the responses were, respectively: 55%, 45%, 0%, and 0%.

Brigham and Wolfskiel's (1983) survey results suggest that attorneys know about some factors that influence **eyewitness identification** accuracy but are less knowledgeable about others. Furthermore, there is little consensus regarding the relative importance of the factors about which they display some knowledge.

Although these two surveys provide insight regarding attorneys' commonsense knowledge of factors that influence **eyewitness identification** accuracy, it is difficult to ascertain from the surveys the extent to which attorneys failed to mention other important factors because of the constraints of the survey or because of their lack of knowledge regarding these factors. Put another way, if attorneys are presented with factors that were not mentioned in the previous studies but that are known to influence **eyewitness identification** accuracy, are they able to identify these potentially harmful factors? It is also impossible, on the basis of surveys, to determine whether attorneys do, in fact, detect the problems they have identified when those problems present themselves in actual cases.

Assessing Attorney Knowledge in Lineup Situations

To address this question and to examine the effectiveness of the presence of counsel at live, postindictment lineups, Stinson et al. (1996) conducted a study designed to assess attorney commonsense knowledge about several factors known to influence lineup suggestiveness, including foil bias (Lindsay & Wells, 1980; Wells & Lindsay, 1980), instruction bias (Cutler, Penrod, & Martens, 1987; Malpass & Devine, 1981; Paley & Geiselman, 1989), and presentation bias (Cutler & Penrod, 1988; Lindsay, Lea, & Fulford, 1991; Lindsay & Wells, 1985).

In the Stinson et al. (1996) study, 109 public defenders were given a photograph of a suspect and were asked to assume the role of the defense attorney for the person in the photograph. The attorneys were then shown one version of a videotaped lineup identification procedure, which depicted a uniformed police office giving either biased or unbiased instructions and presenting either simultaneously or sequentially a foil-biased or foil-unbiased lineup to a female eyewitness.

When presenting biased lineup instructions, the officer simply instructed the witness that she should (a) identify the person she saw commit the crime and (b) not tell other witnesses which lineup member she had identified. When presenting unbiased lineup instructions, the officer read from a card that (a) the lineup might or might not contain the person who committed the crime, (b) the witness should tell him whether or not she saw the person who committed the crime, and (c) she should not tell other witnesses whether or not she had identified anyone.

Lineup presentation was manipulated in the following manner. In the biased lineup presentation, lineup members were presented simultaneously, and the *343 officer simply instructed the witness that she would be seeing a lineup of

five men. In the unbiased lineup presentation, however, lineup members were presented to the witness sequentially, and the officer instructed the witness that (a) she would be seeing a lineup of men who would be presented individually, (b) she had to make a decision after viewing each one, (c) she could not go back to view a lineup member who had already been presented, and (d) she could not be told how many lineup members would be presented.

The lineup members presented to the eyewitness either (a) matched the perpetrator on dimensions of skin tone, height, weight, hair color, and facial hair but varied on other dimensions or (b) matched the perpetrator on no more than two dimensions and thus varied with respect to three of the five dimensions of skin tone, height, weight, hair color, and facial hair.

After viewing the videotaped lineup, the attorneys were asked to assess the suggestiveness and fairness of the foils, instructions, presentation, and overall identification procedure and the extent to which any suggestiveness or unfairness could be corrected. The results indicated that attorneys were sensitive to some factors influencing lineup suggestiveness but appeared to be insensitive to others. Specifically, attorneys were sensitive to foil bias in that they perceived foil-biased lineups as being more suggestive and less fair than foil-unbiased lineups. However, they were only somewhat sensitive to instruction bias; they perceived instruction-biased lineups as being more suggestive but not less fair than instruction-unbiased lineups. Although empirical research has shown that sequentially presented lineups reduce the rate of false identifications (Lindsay, Lea, & Fulford, 1991; Lindsay, Lea, Nosworthy, et al., 1991; Lindsay & Wells, 1985), the attorneys perceived sequential lineups as being significantly more suggestive and less fair than simultaneous lineups.

With respect to lineup correctability, each attorney was asked to rate the likelihood that (a) he or she would submit a motion to suppress the identification based on the suggestiveness of the lineup and (b) the judge would grant such a motion; and, if the identification evidence was presented at trial, that he or she would be able to convince jurors that (c) the identification was inaccurate and (d) the lineup procedure was suggestive. Attorneys who saw foil-biased lineups were more likely to indicate that they would submit a motion, to predict that a judge would rule in favor of the motion, and to believe they would be successful at convincing jurors that the identification was inaccurate and the lineup suggestive. The presence of instruction and presentation biases, however, did not affect attorneys' predictions of whether they would submit a motion or their predictions of the success of the motion and of juror behavior.

In sum, psychological research investigating the effectiveness of the presence of counsel safeguard suggests that attorneys have commonsense knowledge about such factors as race, stress/violence, lighting, viewing conditions, foil bias, and instruction bias but may lack scientific knowledge regarding other factors that influence **eyewitness identification** accuracy, such as eyewitness confidence and presentation bias. In the absence of this scientific knowledge, the presence of counsel safeguard may be less effective than the courts expect. Of course, judges' knowledge of these factors may be identical to attorneys', and their expectations and performance may be commensurately low.

*344 Judges' Commonsense Knowledge About Factors That Influence Eyewitness Performance

Upon identifying some suggestive aspect of an identification procedure, a defense attorney may file a motion to suppress the lineup identification with the court on the grounds that the procedure used to obtain the identification was unduly suggestive. The motion to suppress the identification is then reviewed by the trial judge, who either denies the motion, thereby allowing the identification to be introduced into evidence, or grants the motion, thereby suppressing the identification evidence. Thus, the motion to suppress safeguard rests on the assumption that judges have both commonsense and scientific knowledge about the factors that influence lineup suggestiveness (Stinson, Devenport, Cutler, & Kravitz, 1997).

The research focusing on attorney commonsense knowledge is relevant to judge commonsense knowledge in light of the fact that judges are often experienced attorneys. As mentioned earlier, the research examining attorney commonsense knowledge has demonstrated that attorneys are sensitive to some of the factors that influence **eyewitness**

identification accuracy but are insensitive to others (Brigham & Wolfskeil, 1983; Rahaim & Brodsky, 1982; Stinson et al., 1996).

Stinson et al. (1997) examined judge commonsense knowledge by assessing judges' sensitivity to lineup suggestiveness and the effects of foil, instruction, and presentation biases on judges' rulings on motions to suppress **eyewitness identification.** In this study, 99 Florida judges read a short summary of a hypothetical case involving a robbery and the identification of a suspect by an eyewitness. The case summary provided a written description of both the event and perpetrator as given by the eyewitness, a description of the identification procedure along with a color photocopy of the lineup members, and a standard motion to suppress the lineup identification.

Instruction and presentation biases were manipulated within the description of the identification procedure. Thus, judges read in the description of the identification procedure that (a) the officer had instructed the witness that she should identify the person she saw commit the crime and not tell other witnesses which lineup member she had identified, or (b) the officer had read to the witness that the lineup might or might not contain the person who had committed the crime, that she should tell the officer whether or not she saw the person who committed the crime, and that she should not tell other witnesses whether or not she had identified anyone. Furthermore, judges read that the officer had informed the eyewitness that (a) she would be seeing a lineup of five men, or (b) she would be seeing a lineup of men who would be presented individually, she had to make a decision after viewing each one, she could not go back to view a lineup member who had already been presented, and she could not be told how many lineup members would be presented.

Foil bias was manipulated within the color photocopies of the lineup members. Thus, judges saw either lineup members who (a) matched the perpetrator on dimensions of skin tone, height, weight, hair color, and facial hair but varied on other dimensions or (b) matched the perpetrator on no more than two dimensions and thus varied with respect to three of the five dimensions of skin tone, height, weight, hair color, and facial hair.

*345 After reviewing the case materials, judges were asked to rule on the motion and to rate the suggestiveness and fairness of the foils, instructions, presentation, and overall identification procedure. In sum, judges' rulings on the motion to suppress the identification evidence were influenced by foil and instruction biases. Judges shown foil-biased lineups were more likely to grant the motion and to rate the foils and overall lineup procedure as more suggestive and less fair than judges shown foil-unbiased lineups. In addition, judges who read biased lineup instructions were more likely to grant the motion and to rate the instructions and overall lineup as more suggestive and less fair than judges who read unbiased instructions.

Judges' rulings on the motion to suppress the identification were not affected by presentation bias. Like the attorneys in the Stinson et al. (1996) study, however, judges were more likely to rate sequentially presented lineups as more suggestive and less fair than simultaneously presented lineups. Although empirical research has demonstrated that sequentially presented lineups reduce the rate of false identifications without reducing the rate of correct identifications (see Cutler & Penrod, 1995, for a review), both attorneys and judges appeared to prefer simultaneous lineups and to lack scientific knowledge regarding the beneficial effect of sequential lineups. Attorney and judge preference for simultaneous lineups may be explained by preferences for the underlying differences in the decision-making tasks of simultaneous versus sequential lineups. Specifically, a simultaneous presentation allows the witness to view all lineup members at one time and to compare lineup members, thereby performing a discrimination task. In contrast, sequential presentation requires the witness to perform a more recall-oriented task, of comparing each lineup member solely with his or her memory of the perpetrator rather than to other lineup members (Wells, 1993). Although support for the use of sequentially presented lineups can be found throughout psycho-legal research (Cutler & Penrod, 1988; Lindsay, Lea, & Fulford, 1991; Lindsay & Wells, 1985), 22% of judges sampled criticized the procedure because it did not allow witnesses the opportunity to compare lineup members.

Additionally, Stinson et al. (1996, 1997) suggested that attorney and judge preference for simultaneous lineups may be a result of the fact that attorneys and judges rarely encounter sequentially presented lineups and therefore have

had little exposure to sequential lineup procedures. In fact, a national survey of 220 police departments conducted by Wogalter, Malpass, and Burger (1993) found that sequential lineups are uncommon and are conducted slightly less than 10% of the time.

In summary, research examining judge commonsense knowledge regarding lineup suggestiveness indicates that judges have commonsense knowledge of the harmful effects of foil and instruction biases on **eyewitness identification** accuracy and that their rulings on motions to suppress **eyewitness identification** are influenced by this knowledge. They do not, however, appear to be aware of the scientific research demonstrating the beneficial effect of sequentially presented lineups. On the surface, this research appears to suggest that the motion to suppress the safeguard may be fairly effective, in that judges can accurately evaluate lineup suggestiveness in the form of foil and instruction bias. A small number of judges, however, revealed that they routinely deny these motions, leaving the question of whether a particular identification procedure was suggestive to the jury (Stinson et al., 1997).

*346 Jurors' Commonsense Knowledge About Factors That Influence Eyewitness Performance

If a motion to suppress an identification is denied by the trial judge, the defense attorney is then faced with the task of convincing the jury that the **eyewitness identification** is inaccurate. The defense attorney attempts to carry out this process by emphasizing the suggestive aspects of the identification evidence for the jury through the examination and cross-examination of the identifying witnesses. Thus, the cross-examination safeguard rests on the dual assumptions that (a) attorneys have commonsense knowledge about factors that influence **eyewitness identification** accuracy and thus know what questions to ask during cross-examination, and (b) jurors also have commonsense knowledge about the factors that influence **eyewitness identification** accuracy and use this knowledge appropriately when rendering their verdicts.

Research assessing juror commonsense knowledge of the factors that influence **eyewitness identification** accuracy has taken primarily three approaches. These approaches include (a) surveying juror knowledge, (b) assessing juror ability to predict (actually, postdict) the outcome of **eyewitness identification** experiments, and (c) using mock trials to examine the influence of trial processes such as the cross-examination of witnesses and the manipulation of eyewitness evidence on jury decision making.

Juror Surveys.

Several survey studies have assessed juror commonsense knowledge about the factors that influence **eyewitness identification** accuracy by administering the Knowledge of Eyewitness Behavior Questionnaire (KEBQ) to undergraduates, law students, and community members (Deffenbacher & Loftus, 1982; McConkey & Roche, 1989; Noon & Hollin, 1987). The KEBQ, developed by Deffenbacher and Loftus (1982), contains 14 multiple-choice items involving several **eyewitness identification** scenarios that differ in such aspects as retention interval, training, age of witness, prior photoarray identification, and cross-race identification. Although these studies each tested a different population of respondents from the United States, United Kingdom, and Australia, overall their results were remarkably similar. In general, respondents appeared to be somewhat sensitive to the influence of both cross-race recognition and prior photoarray identification on identification accuracy. However, respondents appeared to be less sensitive to the detrimental effects of age and retention interval. Also, in contrast to the findings of psychological research, respondents believed that training could improve a witness's identification accuracy.

Postdiction Studies of Juror Knowledge.

Juror knowledge of the factors influencing **eyewitness identification** accuracy has also been assessed by having laypersons postdict the outcome of previously conducted **eyewitness identification** experiments (Brigham & Bothwell, 1983; Kassin, 1979, as cited in Wells, 1984). During postdiction studies, students and laypersons read written summaries of an **eyewitness identification** experiment and then postdict the identification accuracy rates of the par-

ticipants in the original experiment. By comparing the postdicted identification accuracy rates with the experimental results, researchers are able to assess the sensitivity of prospective jurors to the specific factors manipulated in the experiment. For example, Kassin (1979) gave students summaries of the experimental conditions in an experiment *347 conducted by Leippe, Wells, and Ostrom (1978). Participants in Leippe et al.'s (1978) study witnessed a staged theft and were led to believe, either before or after the theft, that the theft was high or low in seriousness (i.e., the item stolen was more or less valuable). In the Lieppe et al. experiment, the seriousness manipulation influenced identification accuracy both among eyewitnesses who knew the value of the stolen item prior to the theft (the correct identification rates were 19% in the low seriousness and 56% in the high seriousness conditions) and among eyewitnesses who learned of the stolen item's value after the theft (35% and 13%, respectively). Kassin's participants, however, were not sensitive neither to the influence of crime seriousness on identification accuracy nor to overall levels of identification accuracy; they postdicted the first two cell means would be 66% and 65%, and the second 53% and 60%, respectively.

In one of Wells's (1984) postdiction studies, students read the procedure section of the Leippe et al. (1978) study and were given one of two target cases to predict. In one version, the eyewitness was "completely certain" of his identification; in the other, the eyewitness was "somewhat uncertain." Leippe et al. had found that confidence was not related to identification accuracy. In contrast, Wells's students believed confidence was very strongly related to accuracy: They postdicted a .83 probability of correct identification for the "completely certain" witness and a .28 probability of correct identification for a "somewhat uncertain" witness.

In a second study, Wells (1984) had 80 students read a description of Malpass and Devine's (1981) study of instruction bias, in which participants witnessed a staged act of vandalism. In vandal-absent conditions in the original study, biased instructions produced a 78% false identification rate, versus a 33% false identification rate for unbiased instructions. Wells's students' predictions were 16% and 18%, respectively; they were insensitive to a factor that clearly contributes to the suggestiveness of identification procedures: instruction bias.

Brigham and Bothwell (1983) gave their participants descriptions of two experiments. One was Leippe et al.'s (1978) study of crime seriousness. The second was Brigham et al.'s (1982) field study of cross-race recognition, discussed in the introduction to this article. Brigham and Bothwell found that respondents overestimated the accuracy of **eyewitness identifications**. For the Leippe et al. study, in the condition in which 13% of identifications were actually correct, Brigham and Bothwell's respondents estimated that 71% of identifications would be correct. In the Brigham et al. field study, 32% of White clerks correctly identified Black clerks, but participants in the Brigham and Bothwell study estimated that 51% had done so, and, whereas 31% of Black clerks in the Brigham et al. study correctly identified White customers, Brigham and Bothwell participants estimated that 70% had done so.

In sum, research conducted using postdiction methodology has found that laypersons often predict higher identification accuracy rates than are generally found among participants of eyewitness research (Brigham & Bothwell, 1983; Kassin, 1979, as cited in Wells, 1984). The laypersons in those studies appeared to be insensitive to the influence of crime seriousness (Kassin, 1979), instruction bias (Wells, 1984), and the impact of cross-racial identifications (Brigham & Bothwell, 1983) on **eyewitness identification** accuracy. Furthermore, postdiction studies *348 reveal that laypersons appear to place too much emphasis on eyewitness confidence (Brigham & Bothwell, 1983; Wells, 1984).

Judgment Studies of Juror Knowledge

A third set of studies has examined juror commonsense knowledge by manipulating a number of different factors that have been shown to influence **eyewitness identification** accuracy during simulated trials while other evidence and testimony are held constant (e.g., Bell & Loftus, 1989; Cutler, Penrod, & Dexter, 1990; Cutler, Penrod, & Stuve, 1988; Devenport, Stinson, Cutler, & Kravitz, 1996; Lindsay, Lim, Marando, & Cully, 1986). Participants in this type of study are asked to assume the role of juror while either reading a written summary of a trial, hearing an audiotaped simulation, or viewing a videotaped simulation, after which they are asked to complete questionnaires assessing their

verdicts and other reactions to the trial. Juror sensitivity is gauged by significant effects for factors that are known to influence identification accuracy and nonsignificant effects for factors that are known not to predict identification accuracy (e.g., Cutler, Penrod, & Dexter, 1990).

There have been two distinct approaches to studying juror sensitivity. The first examines jurors' abilities to discriminate between accurate and inaccurate eyewitnesses, and the second examines mock-juror sensitivity to the factors that influence identification accuracy.

An example of the first type of study is an experiment by Wells, Lindsay, and Ferguson (1979) in which the researchers staged a crime in view of witnesses, who then attempted identifications from six-person photoarrays. The researchers then conducted a simulated cross-examination of 24 witnesses who had made accurate identifications and 18 who had made inaccurate identifications. The questions were leading for half of the witnesses and nonleading for the other half. The testimony of these eyewitnesses was then evaluated by 201 undergraduates who served as mock jurors. Wells et al. found that leading questions-typically used in cross-examination-may have a salutary effect on juror assessments of eyewitness performance. When the questions addressed to the witnesses were nonleading, inaccurate eyewitnesses were actually believed by more jurors (86%) than were accurate eyewitnesses (76%). In contrast, when questions were leading, accurate eyewitnesses were believed by more jurors (84%) than were inaccurate eyewitnesses (73%).

Unfortunately, the mock jurors were quite poor at differentiating accurate and inaccurate eyewitnesses. Among jurors exposed to nonleading cross-examination, 76% correctly identified accurate eyewitnesses, but only 14% correctly determined which eyewitnesses were inaccurate. Among jurors exposed to leading cross-examination, 84% correctly identified accurate eyewitnesses, and 27% correctly determined which eyewitnesses were inaccurate. Obviously, the 84% correct classification rate for accurate eyewitnesses is not especially worrisome, but the numbers for inaccurate eyewitnesses are quite disturbing, because they suggest that nearly three out of four mistaken identifications may be believed.

Another worrisome finding of the Wells et al. (1979) study is that eyewitness confidence correlated significantly ($r = .53$) with whether a juror believed the eyewitness, but confidence was not, in fact, significantly correlated ($r = .05$) with the accuracy of this decision. Although jurors were more likely to believe *349 confident eyewitnesses, confident eyewitnesses simply were not more likely to be accurate than less confident eyewitnesses. The proclivity to rely on witness confidence was confirmed in a study by Lindsay, Wells, and Rumpel (1981), who also tested jurors' abilities to discriminate accurate from inaccurate eyewitnesses, using similar methods. Jurors once again gave witness identifications more credence than was appropriate-77% of confident witnesses were believed versus 59% of less confident witnesses. Lindsay et al. also found only a weak relationship between witness confidence and witness accuracy.

Lindsay, Wells, and O'Connor (1989) tested whether mock jurors could differentiate accurate from inaccurate eyewitnesses in a more realistic trial situation. Witnesses to a simulated crime tried to identify the perpetrator from target-present or target-absent photoarrays, and, in a later role play, went through direct examination by a prosecutor, cross-examination by a defense attorney (the attorneys varied in their trial experience), and redirect examination by the prosecutor. These proceedings were videotaped, and 16 "trials" were created with the testimony of 8 eyewitnesses who made correct identifications and 8 who made false identifications. The videotaped trials were then shown to undergraduates, who each viewed one trial and rendered a verdict. The conviction rate did not vary as a function of eyewitness accuracy (i.e., jurors could not differentiate accurate from inaccurate eyewitnesses). The degree of attorney's experience also did not significantly influence verdicts. In short, the realism of the examination and cross-examination, and the experience of the attorneys conducting these examinations, did not change the conclusions from the researchers' earlier studies.

Furthermore, in the largest such study to date, Bothwell, Deffenbacher, and Brigham (1987) used meta-analytical techniques to examine the relation between eyewitness confidence and **eyewitness identification** accuracy. The results of 35 **eyewitness identification** studies were combined, and the relation between witness confidence and

identification accuracy produced an overall estimated effect size of $r = .25$. This effect size suggests that confident eyewitnesses are only somewhat more accurate than unconfident eyewitnesses. Thus, witness confidence level may not prove to be a valuable tool for predicting the accuracy of an **eyewitness identification**.

These studies converge on rather dismaying conclusions about jurors' abilities and indicate that jurors (a) over-estimate the accuracy of identifications, (b) fail to distinguish accurate from inaccurate eyewitnesses, and (c) base their decisions in part on witness confidence-which tends to be a poor predictor of identification accuracy.

To Which Aspects of Eyewitness Identification Are Jurors Sensitive?

Several researchers have attempted to measure jurors' sensitivity to a variety of eyewitness factors-some of which are known from empirical research to influence (or not to influence) eyewitness reliability. In these experiments it was juror sensitivity to these factors that was of interest, rather than ability to differentiate accurate from inaccurate witnesses.

As a first example, Lindsay et al. (1986) conducted a series of four experiments to examine factors that influence jurors' evaluations of eyewitnesses. The first experiment examined the impact of the consistency of identification testimony across eyewitnesses. Lindsay et al. manipulated three factors: (a) strong *350 versus weak physical evidence, (b) the number of eyewitnesses for the prosecution (0, 1, or 2), and (c) the number of eyewitnesses for the defense (0, 1, or 2). The physical evidence did not significantly influence verdicts. The number of eyewitnesses for each side, however, produced significant effects. The overall conviction rates were 10% for 0, 34% for 1, and 45% for 2 prosecution eyewitness, and 41% for 0, 28% for 1, and 21% for 2 defense eyewitnesses. Convictions were most likely when the prosecution's witnesses were unopposed (50%) and least likely when defense's witnesses were unopposed (2%); this finding suggests that the number of eyewitnesses is less important than whether there are conflicting **eyewitness identifications**.

The second Lindsay et al. (1986) experiment examined the effects of defense witness testimony. Five conditions were tested through use of a videotaped enactment of an assault trial: no additional evidence, a second **eyewitness identification** of the defendant, a defense eyewitness who testified that the defendant was not the perpetrator, an alibi witness for the defendant, and an alibi witness who was a relative of the defendant. The highest conviction rates were obtained with two and one unopposed prosecution eyewitnesses (80% and 60%, respectively). When an eyewitness testified that the defendant was not the perpetrator or provided an alibi for the defendant, significantly fewer jurors (27% in each condition) convicted. The alibi provided by the relative did not significantly reduce the conviction rate (57% guilty) in comparison with the no-defense witness conditions.

One factor examined in the third Lindsay et al. (1986) experiment was the impact of inconsistent eyewitness testimony. In one condition there were no inconsistencies in the eyewitness's testimony, but in an inconsistent eyewitness condition, the witness testified that she (a) had originally said the criminal was blond but now did not think the defendant was blond, (b) however did not know whether the defendant had altered her hair color, (c) recalled that the defendant's hair had been dark at the lineup, but (d) was still confident about her identification. The consistency of testimony did not significantly influence jurors' verdicts.

Lindsay et al.'s (1986) fourth experiment examined the impact of viewing conditions at the scene of the crime. In half of the audiotaped trials presented to mock jurors, the crime was said to have occurred at 9 a.m. on a sunny day. In the other half, the crime was said to have occurred at 1 a.m. 60 ft from the nearest streetlight. This factor was crossed with viewing durations of 5 sec, 30 min, or 30 min that included interactions with the perpetrator. In all trials, the eyewitness was highly confident in his identification. Exposure duration's effect on verdicts was nonsignificant. Jurors in the night condition convicted less often than jurors in the day condition (57% vs. 37%), but this difference was also not statistically significant. These results indicate a lack of juror sensitivity to witnessing conditions that influence identification accuracy.

Berman, Narby, and Cutler (1995) also examined the effect of witness consistency on juror evaluations of eyewitnesses, by manipulating whether the witness gave inconsistent testimony on peripheral versus central information. As predicted, jurors exposed to inconsistent eyewitness testimony perceived the eyewitness as less credible and the defendant as less culpable. When exposed to inconsistencies concerning central details, jurors were less likely to convict. The difference between the Berman et al. and the Lindsay et al. (1986) results may stem ***351** from the fact that Berman et al. manipulated inconsistency on more central details than did Lindsay et al.

Bell and Loftus (1989) examined the influence on juror assessments of another aspect of eyewitness testimony-its level of detail. Participants in one of their experiments read a variety of descriptions of a criminal trial. Within these descriptions the degree of detail provided by a prosecution witness varied from high to low, as did the degree of detail provided by a defense witness. The degree of relation between the detail and the perpetrator also varied from high to low. Level of detail was manipulated by using witness statements that contained either the gist of the event or specific details about the event. Relatedness was manipulated by having the statements concern the actions of either the perpetrator or another party. The detail level of testimony influenced participants' verdicts: When the prosecution eyewitness's testimonial detail level was high, 33% of mock jurors convicted, versus 21% when the detail level was low. When the defense eyewitness's testimonial detail level was high, 23% convicted, versus 31% when it was low. These differences were marginally significant. Relatedness did not significantly influence verdicts.

Although exposing inconsistencies in witness testimony and probing about details in the testimony may be effective ways to raise doubts about the reliability of an **eyewitness identification**, it is not clear whether this strategy actually improves the quality of jurors' decisions. Eyewitness studies indicate that description accuracy and consistency are not related to identification accuracy, and therefore jurors probably should not rely on inconsistency as a basis for devaluing eyewitness testimony (see Cutler & Penrod, 1995).

A trial simulation study conducted by Cutler et al. (1988) examined juror commonsense knowledge about 10 factors that vary in the extent to which they influence **eyewitness identification** performance. Participants were 321 students, who each viewed one version of a videotaped trial simulation. Variables included whether the perpetrator had worn a disguise, produced a weapon, or used violence while perpetrating the crime. In addition, the length of the retention interval, the presence or absence of instruction bias and foil bias, and the level of witness confidence were manipulated. The results of this study were consistent with those from both survey and postdiction research. Specifically, when presented with testimony that should call into question the identification accuracy of an eyewitness, jurors appeared to be insensitive to the importance of this information and to pay little attention to it when evaluating the accuracy of the eyewitness. Of the factors studied, jurors relied most on expressions of confidence from the identifying witness ($d = .34$)-although research indicates the relation between witness confidence and identification accuracy is fairly modest (Cutler & Penrod, 1989; Bothwell et al., 1987). A follow-up study conducted by Cutler, Penrod, and Dexter (1990) supplemented the sample with 129 eligible and experienced jurors ($N = 450$) and found that college students and experienced jurors appeared to be comparably insensitive to factors influencing **eyewitness identification** performance.

Finally, a trial simulation conducted by Devenport et al. (1996) assessed juror commonsense knowledge of lineup suggestiveness in the form of foil, instruction, and presentation biases. Participants for this study were 320 students from Nebraska, 160 jury-eligible and experienced citizens from Dade County, Florida, ***352** and 160 jury-eligible and experienced citizens from Lancaster County, Nebraska ($N = 640$) who each watched one version of a simulated videotaped trial. The trial involved a defendant charged with felony murder for the robbery of a convenience store and for the shooting and killing of the store clerk. The primary evidence presented during the trial was the testimony of an eyewitness, who had identified the defendant from a police lineup. Aside from the **eyewitness's identification**, no other evidence incriminating the defendant was presented in the trial.

The eyewitness's testimony focused on the events surrounding the crime and the lineup identification procedure-varying in its descriptions of the lineup instructions given by the police officer (biased or unbiased), the com-

position of the lineup (biased or unbiased), and the manner in which the lineup members were presented (simultaneous or sequential).

The police officer in charge of the investigation testified to the events leading up to the defendant's arrest and also to the procedures used in constructing and administering the lineup to the eyewitness. Specifically, the officer testified to selecting foils for the lineup based on five characteristics of the suspect mentioned by the eyewitness: skin tone, height, weight, hair color, and facial hair. He then testified to (a) selecting men for the lineup if they fit the characteristics mentioned by the witness but varied on characteristics not mentioned by the witness or (b) selecting lineup members who he thought looked like the perpetrator-but who in actuality matched the eyewitness's description of the perpetrator on no more than two characteristics.

The police officer also testified to instructing the eyewitness about the identification procedure prior to showing the witness the lineup. The officer testified that the police department (a) had standardized lineup instructions or (b) did not have standardized lineup instructions, and officers used the instructions they deemed best. The officer then repeated the instructions (biased or unbiased) that he had given to the eyewitness (see Stinson et al., 1996). The instructions also described to the witness the manner in which the lineup members would be presented (simultaneously or sequentially).

After viewing the videotaped trial, jurors were asked to render an individual verdict decision and to rate the overall suggestiveness of the lineup procedure as well as the suggestiveness of the foils, instructions, and presentation of the lineup. The results revealed that, when asked to rate the suggestiveness of the specific lineup biases, jurors were sensitive to evidence of foil and instruction biases but insensitive to the suggestive nature of simultaneous lineups. Specifically, jurors who were presented with evidence of biased foils rated both the foils and the overall identification procedure as more suggestive than jurors who were presented with evidence of unbiased foils. In addition, jurors who were presented with evidence of biased lineup instructions rated the instructions as more suggestive than jurors who were presented with evidence of unbiased lineup instructions. Juror sensitivity to foil and instruction biases, however, did not carry through to influence juror verdicts. These findings suggest that jurors have commonsense knowledge about the harmful effects of both foil and instruction biases, as demonstrated by their awareness of these biases when presented with direct questions about them (which may have the effect of cuing jurors to respond affirmatively), but have difficulty applying this knowledge (or are perhaps unduly influenced by other evidence) in their decision making, as indexed by their *353 verdicts. These results parallel those of the Cutler, Penrod, and Dexter (1990) study, which also found that juror verdicts were not influenced by such factors as use of a disguise, presence of a weapon, violence, retention interval, instruction bias, and foil bias.

It is important to note that the videotaped trial created for this study was designed to present jurors with evidence only relating to an **eyewitness's identification** of a defendant. Throughout the trial, jurors heard testimony regarding the witnessing conditions, which were held constant, and the various procedures used in the lineup identification procedure (foils, instructions, and presentation), which were manipulated. Jurors were not provided with any "hard" evidence such as a gun traced to the defendant or cash linked to the convenience store robbery. They were therefore forced to base their verdict decisions solely on the weight they gave to the identification evidence thereby allowing the researchers to conclude that the jurors were not relying on other case evidence when rendering their verdict decisions. What, then, explains the jurors' ability to identify lineup biases when asked about them directly but their inability to use this information when rendering their verdicts? Although the answer to this question is currently unclear, research examining information load and decision making suggests that when rendering verdict decisions, jurors may have a difficult time using all of the pieces of information that have been presented to them during the course of the trial (Horowitz, ForsterLee, & Brolly, 1996; Malhotra, 1984).

In summary, research conducted with survey, postdiction, and trial simulation methodologies has consistently revealed that jurors tend to rely on factors that are not diagnostic of eyewitness accuracy, such as an eyewitness's memory for peripheral details (Bell & Loftus, 1989) and eyewitness confidence (Wells, 1984), tend to overestimate eyewitness accuracy (Brigham & Bothwell, 1983; Kassin, 1979, as cited in Wells, 1984), and have difficulty applying

their commonsense knowledge of lineup suggestiveness to their verdict decisions (Devenport et al., 1996).

Expert Psychological Testimony and Juror Common Sense

In cases where attorneys believe that the evidence concerning **eyewitness identifications** is outside the common knowledge of the jury, the legal system may permit attorneys to present scientific knowledge, in the form of expert psychological testimony, regarding factors that influence identification accuracy. Although expert psychological testimony on eyewitness memory has been accepted only gradually by the courts (Cutler & Penrod, 1995), this safeguard rests on the assumption that expert psychological testimony regarding the performance of eyewitnesses under various witnessing and identification decisions can increase juror awareness of factors that influence identification performance and assist them in the evaluation of eyewitness evidence. Does this strategy work?

We suggest that using aggregate findings regarding **eyewitness identification** performance to predict the identification accuracy of an individual eyewitness is analogous to predicting whether a prospective car purchase will result in a "lemon." Specifically, any individual eyewitness can be construed as a particular make and model of car manufactured in a particular year. Jurors can be construed as prospective buyers of such cars (**eyewitness identifications**). Expert testimony can be construed as consumer information obtained from a source such as *354 *Consumer Reports* magazine-arguably an expert source regarding product quality.

Car buyers possess a basic level of commonsense knowledge regarding car mechanics and draw upon this knowledge when evaluating a car and predicting whether the car will have mechanical problems. Similarly, jurors have commonsense knowledge regarding eyewitness memory and use this knowledge when evaluating a particular **eyewitness's identification** and when making judgments about whether the witness's identification is accurate. Car buyers can evaluate a car by taking it for a test drive, looking under the hood, and kicking the tires; jurors can predict the accuracy of a witness's identification by examining the witnessing and identification conditions.

In some instances, prospective car buyers may use consumer information sources to supply them with information regarding the general performance of the particular model of car that interests them. Likewise, jurors may be presented with expert testimony regarding factors that can influence an **eyewitness's identification** accuracy. *Consumer Reports* provides the car buyer with information regarding the average performance of the type of car of interest; the scientific knowledge provides jurors with information regarding the average performance of eyewitnesses under various conditions. Although aggregate data regarding the average performance of a specific type of car or eyewitness can assist jurors or car buyers with their decision-making processes, this information cannot ensure that the decision made will be the correct one. (The purchased car may be a lemon, and the believed eyewitness may be wrong.) Aggregate scientific knowledge can, in principle, help jurors and car buyers predict whether a particular **eyewitness's identification** is more or less likely to be accurate and whether a particular car is more or less likely to be a lemon, but, as with all predictions, it does not provide any guarantee that this particular prediction is accurate. Nevertheless, the fact remains the aggregate information can help car buyers and jurors minimize the risk of error.

Studies Involving Expert Testimony

Several studies have examined the influence of expert psychological testimony on eyewitness testimony by manipulating the presence and nature of expert testimony during the trial and then comparing juror conviction rates across conditions (Fox & Walters, 1986; Hosch, Beck, & McIntyre, 1980; Loftus, 1980; Maass, Brigham, & West, 1985; Wells, Lindsay, & Tousignant, 1980). Although early research showed that the presence of expert psychological testimony reduced conviction rates, it was unclear whether this reduction was the result of a simple increase in juror skepticism about eyewitness testimony, as feared by some (McCloskey & Egeth, 1983; McCloskey, Egeth, & McKenna, 1986), or of improved juror sensitivity to the factors influencing **eyewitness identification** performance. *Juror skepticism* refers to a tendency to doubt or disbelieve an eyewitness's testimony, whereas *juror sensitivity* refers to both a general awareness or knowledge regarding the factors that influence eyewitness memory and an ability to use

the information accordingly when rendering a verdict.

By simultaneously manipulating both expert psychological testimony and factors that have been shown to influence witnesses' identification accuracy, a few studies have been able to separate juror skepticism from juror sensitivity and thus *355 determine whether the presence of expert testimony enhances juror sensitivity to the specific factors mentioned during the course of the trial (Cutler, Dexter, & Penrod, 1989; Cutler, Dexter, & Penrod, 1990; Loftus, 1980; Wells et al., 1980). Juror sensitization is evidenced by an interaction between the factors affecting eyewitness memory and expert testimony such that the factors have a larger effect among mock jurors exposed to expert testimony (Cutler & Penrod, 1995).

Loftus (1980) examined both juror skepticism and juror sensitivity in a study assessing the impact of expert testimony on juror verdicts with respect to the level of violence associated with a crime. Mock jurors read a trial transcript in which both the level of violence associated with the crime and the presence of expert testimony were manipulated. The expert testimony described several factors that were related to the case and known to influence **eyewitness identifications**: cross-racial identification, stress, the presence of a weapon, and the use of alcohol. Overall, the results indicated that exposure to expert testimony led to increases in both juror skepticism and juror sensitivity. Thus, participants who were exposed to expert testimony were less likely to convict the defendant (39%) than participants who were not exposed to expert testimony (58%). In addition, jurors appeared to be sensitive to the effect of violence on eyewitness memory. Participants who read a violent version of a trial transcript were more likely to convict (56%) than participants who read a nonviolent version (41%). Furthermore, the results revealed a trend toward enhanced juror sensitivity to the effect of violence when expert testimony was presented during the trial. Specifically, among participants who were not exposed to expert testimony, there was a larger difference in conviction rates between those who read the nonviolent transcript (47%) and those who read the violent transcript (68%) than among participants exposed to expert testimony (35% vs. 43%).

A study conducted by Wells et al. (1980) also assessed both juror skepticism and juror sensitivity to factors such as witness viewing conditions and eyewitness confidence. This study differed from typical trial proceedings in that (a) participants heard only the cross-examination of the eyewitness and (b) participants who were exposed to expert testimony heard the testimony prior to viewing the videotaped cross-examination of the eyewitness. The 108 videotaped eyewitnesses presented in this study had been videotaped for a previous **eyewitness identification** study (Lindsay et al., 1981); they represented **eyewitness identification** conditions in which the **eyewitnesses' identifications** were a result of poor, moderate, or good viewing conditions and included both accurate and inaccurate identifications. Each participant viewed the videotaped cross-examination of four separate eyewitnesses from the same viewing condition and was later asked whether he or she believed the eyewitnesses' testimony. In summary, the results revealed that the presence of expert testimony increased juror skepticism toward the eyewitness testimony. Mock jurors exposed to expert testimony prior to the eyewitnesses' testimony were less likely to believe the eyewitnesses (41%) than mock jurors not exposed to expert testimony (62%). Furthermore, participants were found to be sensitive to the effect of witnessing conditions on identification accuracy, but the trend toward enhanced juror sensitivity by expert testimony was not statistically significant. (The good, moderate, and poor viewing conditions yielded 62%, 58%, and 73% rates of belief in identification accuracy when no *356 expert testimony was presented and 42%, 28%, and 50% rates of belief when expert testimony was presented.)

A study conducted by Cutler et al. (1989) examined the influence of expert testimony on college students' sensitivity to good versus poor witnessing and identification conditions (WIC) and high versus moderate levels of witness confidence. The WIC manipulation involved a combination of factors known to influence **eyewitness identification** accuracy, such as whether the perpetrator had been wearing a disguise, the length of time between the crime and the identification, whether the perpetrator had used a gun, and the instructions presented to the witness prior to viewing the lineup.

The form of expert testimony presented during the trial was also manipulated. When presented, the expert testimony either (a) described how **eyewitness identification** accuracy can be influenced by factors present during

witnessing and identification conditions (descriptive testimony) or (b) provided descriptive testimony as well as information regarding the effect size of specific factors and the number of correct and false identifications obtained under various conditions occurring within **eyewitness identification** experiments (descriptive plus quantified testimony). Additionally, in half of the trials, the expert offered his opinion regarding the accuracy of the identification. The primary dependent measures included in this study were jurors' memory of the trial evidence, jurors' knowledge regarding the factors influencing **eyewitness identification** accuracy, the probability that the identification was correct, and the verdict.

When the data were analyzed with respect to jurors' ratings of whether specific factors influenced identification accuracy, the results revealed that jurors were sensitive to the deleterious effects of disguises, retention interval, and lineup instructions on identification accuracy but were insensitive to the effect of weapon focus and to the trivial relation between eyewitness confidence and identification accuracy. Surprisingly, expert testimony did not enhance juror sensitivity to factors jurors were aware of, but it did increase juror sensitivity to the effects of weapon focus and eyewitness confidence. With respect to verdicts, jurors who were exposed to expert testimony appeared to be more sensitive to the witnessing and identification conditions and more likely to use this information accordingly when rendering a verdict (36% vs. 58% convictions for poor and good WIC, respectively) than jurors who were not exposed to expert testimony (38% vs. 48% convictions for poor and good WIC, respectively). Furthermore, descriptive testimony was found to be a more effective form of expert testimony in terms of improving juror sensitivity to witnessing and identification conditions than descriptive plus qualitative testimony.

The study conducted by Devenport et al. (1996) also assessed the impact of expert psycological testimony on juror commonsense knowledge. By manipulating both lineup suggestiveness and the presence of expert testimony in the simulated videotaped trial, the authors were able to assess the impact of scientific knowledge regarding factors that influence **eyewitness identification** accuracy on jurors' evaluations of lineup suggestiveness.

As described earlier, this study manipulated the presence of foil, instruction, and presentation biases through the testimony of the eyewitness and of the police officer in charge of administering the lineup. Expert psychological testimony, presented by a defense-hired expert, (a) described the encoding, storage, and *357 retrieval stages of memory, and (b) focused on factors known to influence identification accuracy, such as viewing conditions, presence of a weapon, disguises, and factors known to affect lineup suggestiveness. Testimony regarding foil, instruction, and presentation bias was elicited by the defense when a biased identification procedure was used and by the prosecution when an unbiased identification procedure was used.

As mentioned earlier, jurors were sensitive to foil and instruction biases when asked about them directly but were also insensitive to the harmful effects of these factors in rendering their verdicts. More important, expert testimony did not improve juror sensitivity to foil, instruction, or presentation biases with respect to jurors' verdicts, ratings of lineup suggestiveness, and ratings of defendant guilt.

Overall, the research examining the expert testimony safeguard suggests that the presentation of scientific knowledge in the form of expert psychological testimony may enhance juror sensitivity to factors that influence **eyewitness identification** performance such as violence (Loftus, 1980), eyewitness confidence, and other factors present during witnessing and identification conditions (Cutler et al., 1989; Wells et al., 1980). It does not, however, appear to enhance juror commonsense knowledge of factors influencing lineup suggestiveness (Devenport et al., 1996).

Conclusion

The research examining attorney and judge knowledge of factors that influence **eyewitness identification** accuracy is limited. Available research suggests that attorneys are aware of some of the factors influencing the accuracy of **eyewitness identification** evidence but unaware of other factors, such as eyewitness confidence (Rahaim &

Brodsky, 1982) and the presence of a weapon (Brigham & Wolfskiel, 1983). With respect to lineup suggestiveness, attorneys appear to be sensitive to both foil and instruction biases but insensitive to presentation bias (Stinson et al., 1996). Like attorneys, judges are sensitive to the harmful effects of instruction and foil biases, as demonstrated by both their ratings of lineup suggestiveness and their motions to suppress an identification, but are insensitive to presentation bias (Stinson et al., 1997). No research to date has directly assessed judge knowledge or evaluations of **eyewitness identification** evidence with respect to other important factors known to influence **eyewitness identification** accuracy, such as the length of viewing time, presence of a weapon, and prior exposure to mug shots.

The research examining the role of commonsense knowledge in juror evaluations of eyewitness evidence suggests that jurors have commonsense knowledge regarding some factors that influence **eyewitness identification** accuracy but appear to lack scientific knowledge regarding other factors (Bell & Loftus, 1989; Brigham & Bothwell, 1983; Kassin, 1979, as cited in Wells, 1984). Expert psychological testimony appears to improve juror commonsense knowledge about some of the factors affecting **eyewitness identification** memory, such as violence (Loftus, 1980), eyewitness confidence, and other factors influencing witnessing conditions (Cutler et al., 1989; Wells et al., 1980), but it does not appear to enhance juror commonsense knowledge about factors influencing lineup suggestiveness (Devenport et al., 1996).

In sum, the research on attorney, judge, and juror knowledge suggests that the *358 legal safeguards, such as presence of counsel at postindictment, live lineups, motions to suppress identifications, and cross-examination, established to protect defendants from erroneous convictions resulting from mistaken identifications may not be as effective as the legal system intended them to be.

Although there is reason for concern, there are a number of sound policies and practices that offer the prospect of closing the gap between commonsense, legal, and scientific knowledge of factors influencing **eyewitness identification** accuracy. First, the education of court practitioners about the factors that influence eyewitness memory and identification accuracy would not only enhance attorney and judge commonsense and scientific knowledge but also facilitate attorneys' challenges of out-of-court **eyewitness identifications** and judges' decisions regarding motions to suppress identifications, by enabling them to distinguish more readily between biased and unbiased aspects of the lineup identification procedure. Second, given that a survey conducted by Wogalter et al. (1993) indicates that a number of police departments do not use standardized lineup instructions, a nationwide implementation of standardized lineup procedures such as scripted, unbiased instructions that explicitly state that the culprit may or may not be present in the lineup and that give the witness the option to reject the lineup would be an effective and inexpensive way of reducing the rate of false identifications that stem from lineup suggestiveness. Third, as described earlier, more frequent presentation of expert psychological testimony on the factors that influence eyewitness memory and identification accuracy would provide jurors with scientific information that is beyond their commonsense knowledge and improve juror decision making. Finally, there is some evidence suggesting that the use of special judicial instructions that focus on factors known to influence **eyewitness identifications** would also assist jurors with their judgments by providing information that is not within their commonsense knowledge (Greene, 1988; Cutler et al., 1989).

[FNa1]. Note 1. Jennifer L. Devenport and Steven D. Penrod, Law/Psychology Program, University of Nebraska-Lincoln; Brian L. Cutler, Department of Psychology, Florida International University.
    Correspondence concerning this article should be addressed to Jennifer L. Devenport, Law/Psychology Program, University of Nebraska-Lincoln, 209 Burnett Hall, Lincoln, Nebraska 68588-0308.

References

Bell, B. E., & Loftus, E. F. (1989). Trivial persuasion in the courtroom: The power of (a few) minor details. *Journal of Personality and Social Psychology, 56,* 669-679.

Berman, G. L., Narby, D. J., & Cutler, B. L. (1995). Effects of inconsistent eyewitness statements on mock-jurors' evaluations of the eyewitness, perceptions of defendant culpability and verdicts. *Law and Human Behavior, 19,* 79-88.

Bothwell, R. K., Deffenbacher, K. A., & Brigham, J. C. (1987). Correlation of eyewitness accuracy and confidence: Optimality hypothesis revisited. *Journal of Applied Psychology, 72,* 691-695.

Brigham, J. C., & Bothwell, R. K. (1983). The ability of prospective jurors to estimate the accuracy of **eyewitness identifications**. *Law and Human Behavior, 7,* 19-30.

Brigham, J. C., Maass, A., Snyder, L. D., & Spaulding, K. (1982). Accuracy of **eyewitness identifications** in a field setting. *Journal of Personality and Social Psychology, 42,* 673-680.

Brigham, J. C., & Wolfskiel, M. P. (1983). Opinions of attorneys and law enforcement personnel on the accuracy of **eyewitness identifications**. *Law and Human Behavior, 7,* 337-349.

Connors, E., Lundregan, T., Miller, N., & McEwen, T. (1996). *Convicted by juries, exonerated by science: Case studies in the use of DNA evidence to establish innocence after trial* (Report No. NCJ161258). Washington, DC: U.S. Department of Justice, Office of Justice Programs.

*359 Cutler, B. L., Dexter, H. R., & Penrod, S. D. (1989). Expert testimony and jury decision making: An empirical analysis. *Behavioral Sciences and Law, 7,* 215-225.

Cutler, B. L., Dexter, H. R., & Penrod, S. D. (1990). Nonadversarial methods for improving juror sensitivity to eyewitness evidence. *Journal of Applied Social Psychology, 20,* 1197-1207.

Cutler, B. L., & Penrod, S. D. (1988). Improving the reliability of **eyewitness identification**: Lineup construction and presentation. *Journal of Applied Psychology, 73,* 281-290.

Cutler, B. L., & Penrod, S. D. (1989). Forensically relevant moderators of the relation between **eyewitness identification** accuracy and confidence. *Journal of Applied Psychology, 74,* 650-652.

Cutler, B. L., & Penrod, S. D. (1995). *Mistaken **identifications**: The **eyewitness**, psychology, and the law.* New York: Cambridge University Press.

Cutler, B. L., Penrod, S. D., & Dexter, H. R. (1990). Juror sensitivity to **eyewitness identification** evidence. *Law and Human Behavior, 14,* 185-191.

Cutler, B. L., Penrod, S. D., & Martens, T. K. (1987). The reliability of **eyewitness identifications**: The role of system and estimator variables. *Law and Human Behavior, 11,* 223-258.

Cutler, B. L., Penrod, S. D., & Stuve, T. E. (1988). Jury decision making in **eyewitness identification** cases. *Law and Human Behavior, 12,* 41-56.

Deffenbacher, K. A., & Loftus, E. F. (1982). Do jurors share a common understanding concerning eyewitness behavior? *Law and Human Behavior, 6,* 15-30.

Devenport, J. L., Stinson, V., Cutler, B. L., & Kravitz, D. A. (1996). *Does expert psychological testimony improve juror sensitivity to lineup suggestiveness?* Unpublished manuscript.

Finkel, N. J. (1995). *Commonsense justice: Jurors' notions of the law.* Cambridge, MA: Harvard University Press.

Fox, S. G., & Walters, H. A. (1986). The impact of general versus specific expert testimony and eyewitness confidence upon mock juror judgement. *Law and Human Behavior, 10*, 215-228.

Greene, E. (1988). Judge's instruction on eyewitness testimony: Evaluation and revision. *Journal of Applied Social Psychology, 18*, 252-276.

Horowitz, I. A., ForsterLee, L., & Brolly, I. (1996). Effects of trial complexity on decision making. *Journal of Applied Psychology, 81*, 757-768.

Hosch, H. M., Beck, E. L., & McIntyre, P. (1980). Influence of expert testimony regarding eyewitness accuracy on jury decisions. *Law and Human Behavior, 4*, 287-296.

Huff, C. R. (1987). Wrongful conviction: Societal tolerance of injustice. *Research in Social Problems and Public Policy, 4*, 99-115.

Huff, C. R., Rattner, A., & Sagarin, E. (1996). *Convicted but innocent: Wrongful conviction and public policy.* Thousand Oaks, CA: Sage.

Kirby v. Illinois, 406 U.S. 682 (1972).

Krafka, C., & Penrod, S. (1985). Reinstatement of context in a field experiment on **eyewitness identification**. *Journal of Personality and Social Psychology, 49*, 58-69.

Leippe, M. R., Wells, G. L., & Ostrom, T.M. (1978). Crime seriousness as a determinant of accuracy in **eyewitness identification**. *Journal of Applied Psychology, 63*, 345-351.

Lindsay, R. C. L., Lea, J. A., & Fulford, J. A. (1991). Sequential lineup presentation: Technique matters. *Journal of Applied Psychology, 76*, 741-745.

Lindsay, R. C. L., Lea, J. A., Nosworthy, G. J., Fulford, J. A., Hector, J., LeVan, V., & Seabrook, C. (1991). Biased lineups: Sequential presentation reduces the problem. *Journal of Applied Psychology, 76*, 796-802.

Lindsay, R. C. L., Lim, R., Marando, L., & Cully, D. (1986). Mock-juror evaluations of eyewitness testimony: A test of metamemory hypotheses. *Journal of Applied Social Psychology, 16*, 447-459.

*360 Lindsay, R. C. L., & Wells, G. L. (1980). What price justice? Exploring the relationship of lineup fairness to identification accuracy. *Law and Human Behavior, 4*, 303-314.

Lindsay, R. C. L., & Wells, G. L. (1985). Improving **eyewitness identifications** from lineups: Simultaneous versus sequential lineup presentations. *Journal of Applied Psychology, 70*, 556-564.

Lindsay, R. C. L., Wells, G. L., & O'Connor, F.J. (1989). Mock juror belief of accurate and inaccurate eyewitnesses: A replication and extension. *Law and Human Behavior, 13*, 333-339.

Lindsay, R. C. L., Wells, G. L., & Rumpel, C. M. (1981). Can people detect **eyewitness identification** accuracy within and across situations? *Journal of Applied Psychology, 66*, 79-89.

3 PSYPPL 338
3 Psychol. Pub. Pol'y & L. 338

Loftus, E. F. (1980). Impact of expert psychological testimony on the unreliability of **eyewitness identification**. *Journal of Applied Psychology, 65,* 9-15.

Maass, A., Brigham, J. C., & West, S. G. (1985). Testifying on eyewitness reliability: Expert advice is not always persuasive. *Journal of Applied Social Psychology, 15,* 207-229.

Malhotra, N. K. (1984). Reflections on the information overload paradigm in consumer decisionmaking. *Journal of Consumer Research, 10,* 436-440.

Malpass, R. S., & Devine, P. G. (1981). **Eyewitness identification**: Lineup instructions and the absence of the offender. *Journal of Applied Psychology, 66,* 482-489.

McCloskey, M., & Egeth, H. (1983). **Eyewitness identification**: What can a psychologist tell a jury? *American Psychologist, 38,* 550-563.

McCloskey, M., Egeth, H., & McKenna, J. (1986). The experimental psychologist in court: The ethics of expert testimony. *Law and Human Behavior, 10,* 1-13.

McConkey, K. M., & Roche, S. M. (1989). Knowledge of eyewitness memory. *Australian Psychologist, 24,* 377-384.

Noon, E., & Hollin, C. R. (1987). Lay knowledge of eyewitness behaviour: A British survey. *Applied Cognitive Psychology, 1,* 143-153.

Paley, B., & Geiselman, R. E. (1989). The effects of alternative photospread instructions on suspect identification performance. *American Journal of Forensic Psychology, 7,* 3-13.

Penrod, S., & Cutler, B. (1995). Witness confidence and witness accuracy: Assessing their forensic relation. *Psychology, Public Policy, and Law, 1,* 817-845.

People v. McDonald, 37 Cal.3d 351, 690 P.2d 709, 716, 208 Cal.Rptr. 236, 245 (1984).

Pigott, M. A., Brigham, J. C., & Bothwell, R. K. (1990). A field study of the relationship between quality of eyewitnesses' descriptions and identification accuracy. *Journal of Police Science and Administration, 17,* 84-88.

Platz, S. J., & Hosch, H. M. (1988). Cross racial/ethnic **eyewitness identification**: A field study. *Journal of Applied Social Psychology, 18,* 972-984.

Rahaim, G. L., & Brodsky, S. L. (1982). Empirical evidence versus common sense: Juror and lawyer knowledge of eyewitness accuracy. *Law and Psychology Review, 7,* 1-15.

Stinson, V., Devenport, J. L., Cutler, B. L., & Kravitz, D. A. (1996). How effective is the presence-of-counsel safeguard? Attorney perceptions of suggestiveness, fairness, and correctability of biased lineup procedures. *Journal of Applied Psychology, 81,* 64-75.

Stinson, V., Devenport, J. L., Cutler, B. L., & Kravitz, D. A. (1997). How effective is the motion-to-suppress-safeguard? Judges' perceptions of the suggestiveness and fairness of biased lineup procedures. *Journal of Applied Psychology, 82,* 211-220.

United States v. Ash, 413 U.S. 300 (1973).

United States v. Wade, 388 U.S. 218 (1967).

Walters, C. M. (1985). Admission of expert testimony on **eyewitness identification**. *California Law Review, 73*, 1402-1430.

Wells, G. L. (1984). How adequate is human intuition for judging eyewitness testimony. In ***361** G. L. Wells & E. F. Loftus (Eds.), *Eyewitness testimony: Psychological perspectives* (pp. 256-272). New York: Cambridge University Press.

Wells, G. L. (1993). What do we know about **eyewitness identification**? *American Psychologist, 48*, 553-571.

Wells, G. L., & Lindsay, R. C. L. (1980). On estimating the diagnosticity of eyewitness nonidentifications. *Psychological Bulletin, 88*, 776-784.

Wells, G. L., Lindsay, R. C. L., & Ferguson, T. J. (1979). Accuracy, confidence, and juror perceptions in **eyewitness identification**. *Journal of Applied Psychology, 64*, 440-448.

Wells, G. L., Lindsay, R. C. L., & Tousignant, J. P. (1980). Effects of expert psychological advice on human performance in judging the validity of eyewitness testimony. *Law and Human Behavior, 4*, 275-285.

Wogalter, M. S., Malpass, R. S., & Burger, M. A. (1993). How police officers conduct lineups: A national survey. *Proceedings of the Human Factors and Ergonomics Society, 37*, 640-644.

END OF DOCUMENT

A Meta-Analytic Review of the Effects of High Stress on Eyewitness Memory

Kenneth A. Deffenbacher,[1,4] **Brian H. Bornstein,**[2]
Steven D. Penrod,[3] **and E. Kiernan McGorty**[2]

In the past 30 years researchers have examined the impact of heightened stress on the fidelity of eyewitness memory. Meta-analyses were conducted on 27 independent tests of the effects of heightened stress on eyewitness identification of the perpetrator or target person and separately on 36 tests of eyewitness recall of details associated with the crime. There was considerable support for the hypothesis that high levels of stress negatively impact both types of eyewitness memory. Meta-analytic Z-scores, whether unweighted or weighted by sample size, ranged from −5.40 to −6.44 (high stress condition–low stress condition). The overall effect sizes were −.31 for both proportion of correct identifications and accuracy of eyewitness recall. Effect sizes were notably larger for target-present than for target-absent lineups, for eyewitness identification studies than for face recognition studies and for eyewitness studies employing a staged crime than for eyewitness studies employing other means to induce stress.

KEY WORDS: meta-analytic; eyewitness memory; high and low stress.

The performance of eyewitnesses under conditions of heightened stress is of particular forensic interest. When witnessing a crime of violence, the response of the eyewitness is almost always one of generating a stress response to the stressor imposed by the crime. The stress response is actually the defensive response set studied in some detail by psychophysiologists (e.g., Klorman, Weissberg, & Wiesenfeld, 1977). This defensive reaction is the physiological response (acceleration in heart rate, increased blood pressure and muscle tone) that results when the *activation* mode of attention control is dominant (Tucker & Williamson, 1984). The activation mode is one of two neural control systems for regulating response to environmental demands. It is characterized by a tonic readiness for action, a bias against stimulus change, and processing under tight attention controls. Tasks eliciting activation

[1] University of Nebraska, Omaha, Nebraska.
[2] University of Nebraska, Lincoln, Nebraska.
[3] John Jay College of Criminal Justice, New York, New York.
[4] To whom correspondence should be addressed at University of Nebraska, Department of Psychology, 6001 Dodge, Omaha, Nebraska 68182-0274; e-mail: kdeffenbacher@mail.unomaha.edu.

mode dominance include any task serving to increase cognitive anxiety (worry) and/or somatic anxiety (conscious perception of physiological activation), including vigilance, escape, avoidance, or "pressure" tasks (Deffenbacher, 1994).

When assessing any effect of a condition of heightened stress, one must of course compare it with a condition demonstrably lower in stress or even one free of stress, the latter condition being one wherein the *arousal* mode of attention control is dominant (Tucker & Williamson, 1984). Here the physiological response is that of the orienting response (Lacey & Lacey, 1974), a deceleration of heart rate, lowered blood pressure and muscle tone, and an increase in skin conductance with different temporal characteristics than is the case when the activation mode is dominant. The arousal mode of neural control functions to support alert wakefulness and responsiveness to environmental change and novel stimulation. Attention is allocated to the most informative aspect of the stimulus array, rather than being restricted to a specific semantic or motivational content, as is typical when the activation mode of attention control is dominant. Tasks involving simple perceptual intake have been shown to elicit the arousal mode of attention control (Deffenbacher, 1994). Presumably nonthreatening eyewitness events would elicit the arousal mode.

It is clearly important to know just how heightened stress impacts the fidelity of an eyewitness's memory. Since the renaissance of research on eyewitness testimony began in the early 1970s, a scientific literature has accumulated concerning the effects of heightened stress on the fidelity of eyewitness memory. Nevertheless, 30 years of data have not as yet yielded a clear picture of whether heightened stress has a positive, negative, or null effect on eyewitness memory. The principal goal of the present review is to ascertain which of these three possible relations actually obtains.

The first systematic review of the literature relating the effects of heightened stress to eyewitness memory was conducted by Deffenbacher (1983). Typical for the time, Deffenbacher made the assumption that all stressors act to increase general arousal, whether they be high intensity white noise, electric shock, ego-involving instructions, seriousness of a viewed crime, or violence level of a viewed crime. The generally accepted theoretical explanation of the stress-performance relationship was that variations in stressor intensity affected performance level according to an inverted-U function, the function described by the Yerkes–Dodson law (1908). For tasks of at least moderate complexity, and eyewitness identification tasks would appear to qualify, the Yerkes–Dodson law states that performance improves with increases in arousal up to some optimal point and then declines with further increases. In his review of 21 relevant published and unpublished studies, Deffenbacher noted that 10 had produced results which suggested that higher arousal levels increased eyewitness accuracy or at least did not decrease it. The remaining 11 studies produced results showing lowered memory accuracy with increases in arousal. Deffenbacher argued that the studies showing facilitation of memory by arousal increases were likely dealing with arousal increases within the range encompassed by the ascending portion of the inverted-U curve; studies showing memory debilitation with arousal increases were likely operating in the range encompassed by the descending portion of the Yerkes–Dodson curve.

Review of the Effects of High Stress on Eyewitness Memory

A decade later, Christianson (1992) again reviewed the now burgeoning literature relating what he referred to as *emotional stress* and eyewitness memory. He came to rather different conclusions than did Deffenbacher (1983). First, he argued that there was not much evidence to support the notion that emotional stress debilitates eyewitness memory. Second, he proposed that the Yerkes–Dodson law is not an appropriate description of the relation of emotional stress to the fidelity of eyewitness memory. Third, he concluded that in general memory for negative emotional events is better than that for neutral events, at least for central details; typically, however, memory for noncentral details is worse for negative emotional events than for neutral ones. He suggested that the better memory for central details was due to negative emotional events causing greater focusing of attention and increased elaboration of the details within that focus.

Thus far we have established that the studies assessing the effect on memory of what has been variously referred to as heightened stress, anxiety, arousal, or negative emotionality have yielded all possible effects on memory performance, positive, negative, and null. The two reviewers of this literature (Christianson, 1992; Deffenbacher, 1983) arrived at different empirical generalizations that would characterize its body of findings, though admittedly, these two snapshots of the literature occurred a decade apart and included a different mix of research methodologies. Is there a way of resolving this apparent muddle?

We propose that a theoretical alternative to the Yerkes–Dodson law can assist in clearing up at least some of the muddle. In a more recent review, Deffenbacher (1994) revisited the concept of an unidimensional continuum of arousal and concluded that it could no longer be sustained. He likewise concluded that the Yerkes–Dodson law was no longer a useful explanatory construct. Deffenbacher then presented an integrative theoretical alternative to unidimensional arousal theory, a synthesis of Tucker and Williamson's (1984) asymmetric neural control systems model and Fazey and Hardy's (1988) catastrophe model of anxiety and performance, a model that has made some very specific predictions that have been empirically confirmed (e.g., Hardy & Parfitt, 1991). The latter model is a three-dimensional model including two predictor variables, *cognitive anxiety* (worry) and physiological activation, the conscious perception of which has been labeled *somatic anxiety*; the dependent variable is performance. Fazey and Hardy (1988) had concluded from their review of the anxiety-performance literature that any satisfactory model had to be at least three-dimensional. Fazey and Hardy also noted that their model accounted for four different relationships between anxiety and performance found in their literature review. As Deffenbacher (1994) has pointed out, the most interesting prediction from their model is the prediction that at relatively high levels of cognitive anxiety, continuous gradual increases in somatic anxiety (physiological activation) will at first result in continuous, gradual increases in performance, followed at some point by a catastrophic, discontinuous drop in performance. Thus acting in concert, cognitive anxiety and physiological activation produce nonlinear effects on performance. As Deffenbacher (1994) has also noted, a close examination of the data of at least two studies of eyewitness memory (Bothwell, Brigham, & Pigott, 1987; Peters, 1988) confirms the prediction of a catastrophic drop in memory performance at high levels of cognitive anxiety and physiological activation.

Thus by Deffenbacher's (1994) integrative theoretical formulation, if a task elicits the arousal mode of attention control, then memory will be enhanced for the most informative aspects of the stimulus display, those aspects on which the orienting response is focused. If, on the other hand, a task elicits the activation mode of attention control, then memory will either be modestly enhanced or drastically reduced, depending on the relative amounts of cognitive anxiety and physiological activation present.

We are now in a position to begin clearing up the aforementioned muddle of findings. Since 1984, a substantial number of studies have been published showing that increases in what was referred to as *negative emotionality* not only did not adversely impact memory for central details of a scenario but actually improved memory relative to that for central details of a scenario significantly lower in negative emotionality. The only adverse impact on memory by increased negative emotionality was on the less important peripheral details. These studies were a major focus of Christianson's (1992) review and certainly make quite understandable his claim that there was not much evidence to support the notion that emotional stress debilitates eyewitness memory.

We believe that these studies (e.g., Burke, Heuer, & Reisberg, 1992; Christianson, 1984; Christianson, Loftus, Hoffman, & Loftus, 1991; Heuer & Reisberg, 1990; Libkuman, Nichols-Whitehead, Griffith, & Thomas, 1999, Safer, Christianson, Autry, & Osterlund, 1998) were generating facilitation of eyewitness memory for central details, because their principal experimental manipulations likely generated an orienting response (arousal mode of attention control) to stimulating conditions, rather than the defensive response (activation mode of attention control) typically produced by a successful manipulation of stress or anxiety (Deffenbacher, 1994, 1999). The implicit assumption by these investigators appears to have been that higher ratings of negative emotionality for experimental condition stimulus materials (e.g., a modestly gruesome accident or surgery scene) as compared to ratings of control condition stimuli signified a successful manipulation of an emotional state that was akin to stress or anxiety. Another implicit assumption was that by increasing physical exertion not directly relevant to the viewed scenario (e.g., riding an exercise bicycle), a successful manipulation of a physiological state akin to that comprising a defensive response had been attained. As Deffenbacher (1994, 1999) has argued, however, the key experimental manipulation in these studies almost certainly elicited orienting responses, rather than defensive responses.

For one thing, these studies were carried out in laboratory settings where both cognitive anxiety and physiological activation should have been relatively low. Baseline heart rates in these studies averaged 68–82 beats per min (b.p.m.), within the normal range of resting heart rates for young adults. Second, the tasks presented were those of simple perceptual intake or perceptual intake plus instructions emphasizing the need to attend closely to the central and peripheral details of an external event. The negative emotional content of events depicted on key slides was neither a threat to the bodily integrity nor to the self-esteem of the observer. The content was gruesome enough, however, to have elicited an orienting response (Hare, 1972; Hare, Wood, Britain, & Frazelle, 1971; Hare, Wood, Britain, & Shadman, 1970;

Klorman et al., 1977). For example, Hare et al. (1970, 1971) showed that there was not only not a defensive response but in actuality a stronger orienting response to unretouched, color slides of homicide victims than to slides of everyday objects. Hare et al. (1970, 1971) showed heart rate deceleration of 3–5 b.p.m., an important index of an orienting response, as did several of the aforementioned investigators of increases in negative emotionality. Christianson (1984) also noted an increase in skin conductance to stimuli of greater negative emotionality, another index of an orienting response to a stimulus display. Interestingly, Lang, Greenwald, Bradley, and Hamm (1993) found that interest ratings and duration of time an observer chooses to view a visual display both load on the same factor as does the magnitude of the skin conductance response.

A straightforward prediction would be that if a task elicits an orienting response, then memory will be enhanced for the most interesting and informative aspects of the stimulus display, those aspects enjoying the beam of attention provided by the orienting response. Indeed, such memory enhancement was observed in many of the studies involving a manipulation of negative emotionality. Thus, Christianson (1992) was correct that the better memory for central details in these studies was due to a greater focusing of attention on them. However, the relevant mechanism was a qualitatively different one than the one he supposed, the orienting response of the arousal mode of attention regulation, rather than the defensive response characteristic of the activation mode of attention regulation. Hence the results of these studies are not relevant to assessing the effect of heightened stress on the fidelity of eyewitness memory.

Thus, we are limiting our focus to studies whose experimental manipulations were actually productive of a difference in stress response level. Now given that neither the seminal Shapiro and Penrod (1986) meta-analysis of the eyewitness literature nor any subsequent one has addressed the effect of heightened stress on eyewitness memory, it would be desirable to have a firm estimate of effect size and direction, both in regard to accuracy of face identification and recall of details. As mentioned earlier, this is the primary goal of the present review. Other goals include identifying variables that might moderate any consistent effect of stress on the fidelity of memory, identifying any methodological or theoretical shortcomings in the body of relevant literature, and to consider possible directions for future theoretical development and research.

METHOD

Sample

Inasmuch as the present review was part of a comprehensive meta-analysis project, a thorough search of social science citation retrieval systems was conducted. These systems included PsycINfO, Educational Resources Information Center (ERIC), *Sociological Abstracts*, *Dissertation Abstracts International*, Dissertations on-line (http://www.contentville.com/content/dissertations.asp), Medline, and Social Scisearch (the *Social Science Citation Index*). These computer database searches were supplemented with more traditional search methods, including use of

bibliographic citations in published research and in social science pro- ceedings and contacting leading researchers, in order to identify the most recent published research.

No unpublished studies were included, because the legal standards for prof- fered scientific testimony established by the US Supreme Court in *Daubert v. Merrell Dow Pharmaceuticals* (1993) have strengthened the preference by the legal system for meta-analytic conclusions based on a body of well conceived, well exe- cuted, and easily retrievable studies. In order to be included, a published study must have met three criteria: (a) a statistical test of the effect of heightened stress or anx- iety on one or more measures of eyewitness memory accuracy must have been pro- vided; (b) either stress/anxiety must have been manipulated directly or have been included as a quasi-experimental independent variable; and (c) manipulation checks must have been provided showing that putative manipulations of stress/anxiety or perceived level of violence had been successful.

There were in addition three specific exclusion criteria for studies wherein the claim was made that the relation between stress and eyewitness memory had been successfully tested. First, ratings of witness stress or anxiety, whether self-ratings or by others, had to be concurrent, as soon after encoding of the target person(s) as possible, not retrospective. Some studies of children's memory for medical proce- dures have involved retrospective ratings that were delayed by periods of a week up to as much as a year or more (e.g., Peterson & Bell, 1996; Quas et al., 1999). Second, measures of memory had to be from an initial assessment, a measure of memory unsullied by previous attempts at identification or recall. Again, some studies of children's memory for medical procedures have focused on long-term recall after two or more previous assessments of recall accuracy (e.g., Burgwyn-Bailes, Baker- Ward, Gordon, & Ornstein, 2001; Peterson & Whalen, 2001). Third, measures of recall accuracy that were included were only of straightforward efforts at either free recall or interrogative recall (cued recall). Measures of recall accuracy after attempts at misleading postevent suggestion were not included (e.g., Bruck, Ceci, Francoeur, & Barr, 1995).

There were two final study samples. The first sample included 16 published papers, providing a total of 27 independent estimates of effect size for heightened stress on accuracy of *face identification*. This sample included work published be- tween 1974 and 1997, with a total of 1727 participants involved in relevant tests of the stress/anxiety effect. Sample sizes across the 27 tests of the effect ranged from 18 to 165 ($M = 64.0$).

In order to conduct a companion meta-analysis of the effects of heightened stress on accuracy of *eyewitness recall* (of perpetrator characteristics, crime scene details, and actions of central characters), another sample of 18 published pa- pers meeting the aforementioned criteria for inclusion was collected. Ten of these published papers were also included in the prior sample, studies which had in- cluded measures both of face identification accuracy and accuracy of recall. This latter sample provided 36 independent estimates of effect size and yielded a to- tal of 1946 participants in the various tests of the effect of stress/anxiety on accu- racy of eyewitness recall. Across the 36 tests, sample sizes varied from 18 to 249 ($M = 54.1$).

Review of the Effects of High Stress on Eyewitness Memory

Study Characteristics

The previously mentioned comprehensive meta-analysis is intended to update and to extend the one conducted by Shapiro and Penrod (1986). The current comprehensive meta-analysis, from which the current study springs, ultimately has encompassed coding of approximately 450 existing studies of face recognition from both the eyewitness and laboratory face recognition memory traditions. Several dozen independent variables were coded for each of these studies. These variables included stable (e.g., sex and race) and malleable (e.g., disguise) characteristics of both participants and targets, situational (e.g., exposure duration), and procedural (e.g., lineup presentation) factors. Variables worthy of specific mention for their usefulness in the present review included type of study (eyewitness identification study or laboratory face recognition study), whether or not the study employed a staged crime, whether the lineup included the target-present (TP) or target-absent (TA), number of participants, age of participants, and most important, whether anxiety, stress, or violence level was manipulated.[5] Dependent variables recorded were proportion correct, hit and false alarm rates, if provided, for TP lineups, correct rejection and false alarm rates, when provided, for TA lineups, and the signal detection measures, d-prime and beta.

Statistics

In order to test the statistical reliability of an estimate of the typical effect size found in any particular meta-analysis, we have adopted the Stouffer method (Rosenthal, 1995). Here a meta-analytic $Z(Z_{ma})$ was calculated by combining Z-scores associated with individual tests of the hypothesis that heightened stress negatively impacts eyewitness memory. The resulting algebraic sum, when divided by $\sqrt{k}$, where k is the number of independent estimates of the effect size, yields the meta-analytic Z. The probability associated with the meta-analytic Z is the overall probability of a Type I error associated with the observed pattern of results. Inasmuch as Z_{ma} provides an unweighted estimate of the overall probability level, a meta-analytic $Z(Z_{mn})$ was also calculated which weighted individual Z-scores by sample size of the study; this allows estimation of population parameters with greater emphasis on larger samples and their more reliable parameter estimates.

It should be noted that whenever recovery of sample sizes and proportion of correct identifications per condition permitted, the Z-score entered into the meta-analysis was one calculated for the difference between proportions. When an exact Z-score could not be calculated for a given effect size estimate, a Z-score associated with the p value for the estimate was entered, 1.65 for $p = .05$, for instance. When a test of the hypothesis was reported as not significant, but no statistics were cited, the conservative procedure of entering $Z = .00$ was followed (Rosenthal, 1995).

---

[5]To assess the reliability of coding study independent variables, two raters generated separate codings for each of 80 variables across a randomly selected 50% of the studies included in the present meta-analysis. Rate of agreement across all variables and 14 studies averaged 93%.

App-47

Deffenbacher, Bornstein, Penrod, and McGorty

All studywise differences between proportions for high and low stress conditions were converted to the effect size h, inasmuch as h is the coefficient recommended by Cohen (1988) when testing for differences between proportions. Hence, mean effect size for any set of studies and its associated 95% confidence interval is expressed in terms of h.

Finally, given that we have included only published studies in the present meta-analysis, it is quite clear that our sample of studies is not a random sample of all studies that may actually have been conducted. As Rosenthal (1995) has pointed out, it is rather likely that published studies have reported lower probabilities of a Type I error than have those studies "squirreled away in file drawers." The concern in regards to this "file drawer problem" is that a sufficient number of such studies averaging null results could threaten a meta-analytic conclusion. We have therefore employed Rosenthal's (1995) suggested procedure for calculating a fail-safe $N(N_{fs})$ in order to determine the number of unknown or not retrieved studies averaging null results required to increase the probability of a Type I error to the just significant level of $p = .05$. Actually, inasmuch as this number is typically a whole number plus a fractional number of studies, we have adopted the rule of rounding to the next higher number. Thus, most values of the fail-safe N that we report in connection with a meta-analytic Z, represent the number of additional null results studies required to increase our probability of a Type I error to a value slightly greater than .05. Clearly, the fail-safe N represents a "tolerance for future null results" (Rosenthal, 1995). We would propose that at an absolute minimum the fail-safe N must be at least as large as the number of independent estimates of effect size that went into calculating the meta-analytic Z.

RESULTS

Meta-Analysis 1: Identification Accuracy

All Tests

We first sought to determine the overall status of the hypothesis that heightened stress debilitates eyewitness memory for faces. For this analysis proportion correct for the low stress condition was subtracted from that of the high stress condition. This has the virtue of producing a positive meta-analytic Z should high stress facilitate eyewitness memory and a negative value should high stress debilitate eyewitness memory. In this instance, overall proportion of correct identifications for the high stress condition was .42; for the low stress condition, it was .54. Mean effect size, h, for this analysis was $-.31$ (95% *CI*: $-.04$ to $-.58$); median effect size was $-.27$. The meta-analytic Z (Z_{ma}) was -6.44, $p < .0001$, $N_{fs} = 390$. Weighting each of the 27 independent tests of the hypothesis by sample size yielded $Z_{mn} = -6.03$, $p < .0001$, $N_{fs} = 336$ studies. These analyses provide clear support for the hypothesis that heightened stress has a negative impact on eyewitness identification accuracy. In subsequent analyses, we tested for moderator variables which might account for the considerable variability in effect sizes of individual tests of the hypothesis (effect size range: -3.02 to $+ .52$; $s = .68$; see Table 1). Table 2 contains a summary of all effect sizes associated with Meta-Analyses 1 and 2.

Review of the Effects of High Stress on Eyewitness Memory

Table 1. Identification Accuracy Effect Sizes (Studies Ordered by Date)

Author	Date	N	Overall (h)	TP (h)	TA (h)
Buckhout et al.	1974	48	−3.02	−3.02	
Mueller et al.	1979	96	.04	.04	
Nowicki et al., Exp. 1	1979	93	−.35	−.35	
Bailis & Mueller	1981	120	.13	.13	
Clifford & Hollin	1981	60	−.16	−.16	
Brigham et al.	1983	20	−.46	−.46	
Bothwell et al., Neurotics[a]	1987	35	−.74		
Bothwell et al., Stables[a]	1987	36	.52		
Cutler et al.[a]	1987	165	−.08		
Tooley, Brigham, Maas, and Bothwell	1987	96	.14	.14	
Peters, TP	1988	106	−.51	−.51	
Peters, TA	1988	106	−.18		−.18
Maas & Kohnken[b]	1989	86	−.43		−.43
Hosch & Bothwell, Exp. 1[a]	1990	39	.49		
Kramer et al., Exp. 1[b]	1990	64	−.58	−.58	
Goodman et al., Exp. 1	1991	18	−.47	−.47	
Goodman et al., Exp. 2	1991	47	−.27	−.27	
Goodman et al., Exp. 3	1991	34	.24	.24	
Peters, Exp. 1, TP	1991	36	−.57	−.57	
Peters, Exp. 1, TA	1991	35	−.04		−.04
Peters, Exp. 2, TP	1991	34	−.61	−.61	
Peters, Exp. 2, TA	1991	33	.32		.32
Peters, Exp. 3, TP	1991	32	−1.32	−1.32	
Peters, Exp. 3, TA	1991	32	.39		.39
Peters, Exp. 4[a]	1991	96	−.12		
Peters, Exp. 1[a]	1997	64	−.42		
Peters, Exp. 2[a]	1997	96	−.31		

[a]Only overall proportion correct reported.
[b]Weapon visibility totally confounded with anxiety level; thus reportedas a stress effect.

Lineup Type

Because of a comment by Peters (1988) that he had not found the same statistically reliable difference between high and low stress conditions for TA lineups that he had found for TP lineups and because of the forensic implications of TA and TP lineups, we decided to code both conditions as independent estimates of the effect of heightened stress on eyewitness identification for all those studies manipulating lineup type as a between subjects variable (see Table 1). Lineup type clearly emerged as the most powerful moderator of the impact of stress on face identification accuracy (hits and correct rejections). TP lineups ($N = 15$) generated a mean effect size $h = -.52$, 95% CI: −.08 to −.96. Here $Z_{ma} = -7.08$, $p < .0001$, and $N_{fs} = 264$. On the other hand, TA lineups ($N = 5$) generated a negligible mean effect size, $h = +.01$ (95% CI: −.39 to + .41). In this case, $Z_{ma} = -.56$, which was clearly not statistically reliable. Yet another way to examine the different impact of heightened stress on face identification accuracy as a function of lineup type is to note that mean proportions correct for TP lineups under high and low stress conditions were .39 and .59, respectively. Corresponding mean proportions correct for TA lineups were .34 and .34. Thus, the overall negative impact of heightened stress on accuracy of face identification was due entirely to a substantial effect on hit rate for TP lineups. The correct rejection rate for TA lineups was unaffected by stress level.

Table 2. Meta-Analysis Effect Sizes

	Effect size	
Type of Analysis	h	d
Identification accuracy		
All tests[a]	−.31	
All tests[b]	+ .22	
TP lineups[a]	−.52	
TA lineups[a]	+ .01	
TP lineups[b]	+.37	
TA lineups[b]	.00	
Identification paradigm[a]	−.36	
Recognition paradigm[a]	−.10	
Staged crimes[a]	−.58	
Other stressors[a]	−.28	
Adult witnesses[a]	−.34	
Child witnesses[a]	−.27	
Recall accuracy		
All tests		−.31
Interrogative recall		−.34
Narrative recall		−.20
Adult witnesses		−.44
Child witnesses		−.06
Staged crimes		−.45
Other stressors		−.16

[a]Overall proportion correct, including hit and correct rejection rates.
[b]False alarm rate.

Lineup type moderated the effects of heightened stress on the rate at which faces were falsely identified, as well. In this instance, mean false alarm rates for TP lineups under high and low stress conditions were .34 and .19, respectively. Comparable false alarm rates for TA lineups were .66 and .65. Accordingly, it is not surprising that TP lineups $(N = 7)$ generated a mean effect size $h = +$ .37, 95% CI: + .05 to + .69. Though $Z_{ma} = +3.36$, $p = .0004$, the fail-safe N was only 23 additional null-results studies. In parallel with the results for proportion correct, the false alarm rates for TA lineups $(N = 5)$ were not differentially affected by stress levels, with the result that mean effect size was negligible, $h = .00$ (95% CI: −.43 to +.43). The meta-analytic Z_{ma} was + .56, certainly not significant. Clearly, the overall tendency of heightened stress to increase false alarm rates in a face identification task, $h = +.22$ (95% CI: −.02 to +.46), $Z_{ma} = +2.92$, $p = .0018$, $N_{fs} = 26$, was due entirely to a substantial effect for TP lineups.

Research Paradigm

A somewhat smaller, though still sizeable moderator effect was found for the variable of research paradigm, whether the study was conducted as a standard face recognition task in the tradition of cognitive psychology or whether it was conducted in the context of the eyewitness identification paradigm. The face recognition task exposes observers to a relatively large number of target faces (at least 24 in

Review of the Effects of High Stress on Eyewitness Memory

studies included in the present meta-analysis). A recognition memory test following exposure to the target faces usually includes twice as many faces, the targets plus an equal number of unfamiliar distracter faces. Observers are exposed to the test faces one at a time and are instructed to respond "yes" or "no" as to whether a given face had been exposed previously. Studies conducted in the eyewitness identification paradigm usually expose witnesses to just one or two target faces, the perpetrator(s), and memory for each target's face is tested either by embedding his/her face in a 5–9-person simultaneously or serially presented live lineup or photo spread (TP lineup) or else by substituting someone else who is a match to the perpetrator's description (TA lineup). Witnesses are asked to identify the perpetrator or to indicate that he/she is not in the lineup.

For the face recognition studies included in our sample ($N = 5$) the mean proportion correct under high stress conditions was .56 and was .58 under low stress conditions. Not surprisingly, the mean effect size in this instance was only $-.10$ (95% CI: $-.45$ to $+.25$). Even though the meta-analytic Z was significant $Z_{ma} = -2.46$, $p = .0069$, the fail-safe N was only seven additional null results studies.

Mean proportions correct were .39 and .53 under high and low stress conditions, respectively, for witnesses in the 22 studies executed in the more ecologically valid eyewitness identification tradition. This difference resulted in a mean effect size $h = -.36, 95\%$ CI: $-.04$ to $-.68$. The debilitating effect of heightened stress on eyewitness memory for studies conducted in the eyewitness identification paradigm was a statistically reliable one, $Z_{ma} = -6.00, p < .0001$, a conclusion not likely to be overturned by unknown null results studies, $N_{fs} = 269$. Clearly, heightened stress is much more likely to have a debilitating effect on memory for the human face when encoding and memory testing occur under the requirements of the eyewitness identification paradigm than when encoding and testing occur under conditions of the face recognition memory paradigm.

Presence/Absence of a Staged Crime

Within the 22 eyewitness identification studies, six manipulated stress in the context of a staged crime, and 16 manipulated stress by some other means, threat of an injection, for instance. Mean proportions correct under high and low stress conditions were .33 and .50, respectively, for the staged-crime studies. For the studies manipulating stress by other means, the comparable means were .56 and .69. Even though the adverse effect of heightened stress on eyewitness memory was statistically reliable for both sets of studies, $Z_{ma} = -3.82, p < .0001, N_{fs} = 27$ for the staged-crime studies and $Z_{ma} = -4.68, p < .0001, N_{fs} = 113$ for the other studies, there was a pronounced difference in mean effect sizes generated. Mean effect sizes were $h = -.58, 95\%$ CI: -1.88 to $+ .72$, for the staged crime studies and for the studies manipulating stress by other means, a smaller $h = -.28, 95\%$ CI: $-.02$ to $-.54$. One study (Buckhout, Alper, Chern, Silverberg, & Slomovits, 1974) was responsible for most of this difference in effect sizes, however. Nevertheless, it should be noted that the study of Buckhout et al. (1974) was a rather realistic, *live* staged crime, rather than a filmed one.

Deffenbacher, Bornstein, Penrod, and McGorty

Age

Given that 15 of our independent estimates of the effect size of heightened stress on face identification were produced by adult witnesses/observers and that 12 were produced by children (ages ranged from 3 to 10 years), we decided to test age as a moderator of effect size. It was also of interest to assess age as a possible moderator, given the concern about competency of child witnesses (e.g., Goodman, Hirschman, Hepps, & Rudy, 1991). At least for face identification accuracy, witness age appears to have contributed little to the variability in effect size. Mean proportions correct for children under high and low stress conditions were .42 and .55, respectively, while for adults the proportions were .42 and .54. These differences generated average effect sizes that were comparable, $h = -.27$, 95% CI: $-.57$ to $+ .03$, for the children and $h = -.34$, 95% CI: $-.80$ to $+ .12$, for the adults. The debilitating effects of heightened stress on identification accuracy were statistically reliable in both instances. For the children, $Z_{ma} = -3.43, p < .0003, N_{fs} = 41$; for the adults, $Z_{ma} = -5.61, p < .0001, N_{fs} = 159$ studies.

Meta-Analysis 2: Accuracy of Eyewitness Recall

All Tests

We should first note that it was not possible to calculate the effect size h across all 36 tests of the hypothesis that heightened stress debilitates eyewitness recall. In only five of these instances did investigators report proportion of details correctly recalled as a function of stress level; for these five studies $h = -.25$, corresponding to an average proportion correctly recalled of .52 in the high stress condition and .64 in the low stress condition. Consequently, d was adopted as a substitute measure of effect size. Again, Table 2 summarizes all effect sizes reported for Meta-Analysis 2.

Both meta-analytic Zs were statistically significant, $Z_{ma} = -5.40, p < .0001, N_{fs} = 355$ studies, and $Z_{mn} = -6.06, p < .0001, N_{fs} = 453$ studies. Calculation of the mean effect size yielded $d = -.31$, 95% CI: $-.14$ to $-.48$. Clearly, heightened stress produces the same debilitating effect on accuracy of eyewitness recall as it does on identification accuracy. In the remaining analyses, we tested for variables that might have moderated this effect on recall.

Type of Recall

Eight estimates of effect size were associated with narrative recall (free recall); the remaining 28 estimates involved some form of interrogative recall (specific questions). The meta-analytic Z for narrative recall was not statistically reliable, $Z_{ma} = -1.17$. Average effect size was $d = -.20$ (95% CI: $-.68$ to $+ .28$), all the effect being generated by a single study (Clifford & Scott, 1978). For interrogative recall, $Z_{ma} = -5.50, p < .0001, N_{fs} = 288$. Here d was $-.34$, with a 95% CI extending from $-.15$ to $-.53$. Heightened stress would appear to impact interrogative recall much more negatively than narrative or free recall.

Age

In this sample of effect size estimates, there were 23 tests with adult witnesses and 13 tests with children as witnesses (age range again 3–10 years). Unlike the situation with identification accuracy, here age emerged as an important moderator. For adults, $d = -.44$ (95% *CI*: $-.19$ to $-.69$), while for children, d was a negligible value, $-.06$ (95% *CI*: $-.16$ to $+ .04$). There was a significant meta-analytic Z for adult eyewitnesses, $Z_{ma} = -6.05, p < .0001, N_{fs} = 286$ studies. For child eyewitnesses, on the other hand, $Z_{ma} = -1.01$, a not statistically reliable value. Surprisingly enough, heightened stress debilitated eyewitness recall for adults, but not for children. However, before concluding that the null hypothesis that $Z_{ma} = 0.00$ might have validity for children, one should consider that in this instance the *counternull hypothesis* (Rosenthal, 1995), $Z_{ma} = -2.02$, is just as likely to be true.[6] Even were the counternull hypothesis true, however, it would still be the case that witness age is an important moderator of the effect of heightened stress on eyewitness recall.

Presence/Absence of a Staged Crime

Though the distinction between face recognition and eyewitness identification research paradigms is not applicable to eyewitness recall, the distinction between presence and absence of a staged crime is indeed relevant. In our sample of studies assessing eyewitness recall as a function of stress level, there were 18 independent estimates of effect size that included a staged crime, on film or live, and 18 that did not. Meta-analytic Zs were statistically significant in both instances, $Z_{ma} = -4.67$, $p < .0001, N_{fs} = 127$ for the studies including a staged crime and $Z_{ma} = -3.01, p = .0013, N_{fs} = 42$ for those investigations not including a staged crime. Despite both conditions producing statistically reliable decrements in recall under higher levels of stress, the presence of a staged crime would appear to have generated a somewhat greater decrement. In support of this assertion, we may note that the effect size generated by the staged crime studies ($d = -.45$; 95% *CI*: $-.17$ to $-.73$) was more than twice that generated by the studies employing other means to induce stress ($d = -.16$; 95% *CI*: $-.35$ to $+ .03$).

DISCUSSION

By adopting our particular inclusion criteria for our two samples of studies, we sought to limit our focus to experimental manipulations productive of defensive responses to stimulating conditions. In so doing, we have adduced considerable support for the hypothesis that high levels of stress negatively impact both accuracy of eyewitness identification as well as accuracy of recall of crime-related details. For eyewitness identification, the average effect size h was $-.31$, with a 95% *CI* that did not include zero. Whether unweighted or weighted by sample size, the meta-analytic Zs were associated with fail-safe Ns of 300–400 studies. Thus, the current

---

[6]Here the counternull value of the effect size of heightened stress is found by doubling the obtained effect size ($Z_{ma} = -1.01$) and subtracting the effect size expected under the null hypothesis, .00.

meta-analytic conclusion regarding the negative effect of stress on eyewitness identification accuracy is unlikely to be overturned any time soon by unknown findings averaging null results. The conclusion that heightened stress debilitates eyewitness recall (average effect size $d = -.31$), too, is at least as safe, with fail-safe Ns of more than 350 studies associated with the overall meta-analytic Zs.

In addition to ascertaining the direction and magnitude of the effect of heightened stress on accuracy of eyewitness memory, we had sought to identify variables that might moderate any consistent effects of stress on the fidelity of memory. We found two principal moderators of the negative effect of heightened stress on eyewitness identification accuracy, lineup type and research paradigm. Certainly there is precedent for the importance of lineup type as a moderator of accuracy in meta-analyses of the eyewitness identification literature (e.g., Steblay, 1997). Steblay found a moderate size effect for unbiased instructions to increase accuracy in TA lineups but to have no effect on accuracy in TP lineups. The moderator effect for lineup type found here was the reverse of the one found by Steblay. Here TP lineups generated an effect size ($h = -.52$), such that face identification accuracy was much reduced under conditions of heightened stress as compared with low stress conditions, a difference of .20 in mean proportion correct/hit rate (.39 versus .59, respectively).[7]

Quite possibly, encoding of a target person under conditions of heightened stress reduces the veridicality of a witness's memory representation of him sufficiently to decrease the probability of a match between that representation and the target when present (TP lineups). Such an effect of stress would certainly serve to reduce the hit rate. The proportion correct measure for TA lineups, the correct rejection rate, might be similarly affected, inasmuch as reduced quality of a witness's memory representation of the target would not provide as good a basis for rejecting a lineup that does not contain his face. On the other hand, members of a fair TA lineup only roughly resemble the target person; their faces are consistent with just a witness's prior verbal description of the target, not necessarily with a high-quality *visual* representation of him. Perhaps, therefore, differences between the relatively nondegraded and stress-degraded visual memory representations of the target are insufficient to affect the basis for deciding that the target face is not present in the TA lineup. Indeed there was a negligible effect size of heightened stress on accuracy in TA lineups; proportion correct/correct rejection rates were .34, regardless of stress level. Inasmuch as correct rejection rates and false alarm rates for TA lineups must sum to 1.00, there would likewise be no difference expected in false alarm rates for TA lineups as a function of differences in stress at encoding, .66 in both cases.

We should note in passing that lineup type moderated the effect of heightened stress on the false alarm rate, as well. TP lineups generated an effect size ($h = +.37$) that was considerably larger than the negligible effect generated by TA lineups ($h = .00$). This TP effect corresponded to false alarm rates of .34 under high

---

[7]Another and perhaps more meaningful way to interpret a difference in proportion correct of this magnitude has been suggested by G. L. Wells (cited in Steblay, 1997). Consider, for instance, 1000 TP lineups conducted over a period of time. Given the accuracy difference we have obtained, we would expect TP lineups to generate 200 more correct identifications of perpetrators witnessed under low stress conditions than of perpetrators witnessed under high stress conditions.

stress conditions and .19 under low stress conditions. The aforementioned stress-induced memory degradation could account for the increased false alarm rate, as well, if indirectly. Having reduced the probability of a match to the target in a TP lineup, the stress-induced loss of memory fidelity would then serve to increase the probability of a "match" to some other lineup member, in direct proportion to the motivation of the eyewitness to choose someone from the lineup. In the present instance, much of the stress-induced lowering of the hit rate (.20) was likely "transferred" to an increase in the false alarm rate (.15); the remaining decrease in the hit rate (.05) might have resulted in an increase in the false rejection rate for a TP lineup, the latter rate rarely reported in the literature. Obviously, in a TA lineup there is no possibility of a similar transfer of witness choices from one person, the target, to other possible choices. At any event, our results portend that a greater proportion of high-stress witnesses than low-stress witnesses will choose a foil from a TP lineup and will thus have their subsequent credibility as a witness undermined.

We should also note that even though the stress effect resides in TP arrays, plausible arguments can be made that the mix of guilty and innocent persons identified by witnesses will change as a consequence, even if the proportion of positive identifications from TA lineups does not. For simplicity's sake, let us assume that of all lineups conducted by police, half are TP and half are TA. Given the .59 hit rate in low-stress TP lineups and the .66 false alarm rate obtained in our sample of TA lineups, this means that 59 guilty perpetrators would be identified from every 100 TP lineups and 66 mistaken identifications would be made from every 100 TA lineups. If we were also to assume that all lineups were perfectly fair 6-person arrays, an innocent suspect embedded in a TA lineup would be mistakenly identified 11 times (66/6) per 100 arrays. Hence, the resulting pool of identifications would be 59 guilty and 11 innocent, an accuracy rate of .84 for choosers of suspects—of course, we should not forget that 41 guilty perpetrators would not be identified. Extending our argument to the high-stress situation, our results imply that 100 TP arrays would generate identifications of 39 guilty perpetrators and that 100 TA arrays would generate 66 mistaken identifications. An innocent suspect in the TA lineups would again be chosen 11 times; the resulting pool of identifications would be 39 guilty and 11 innocent, an accuracy rate of .78 for choosers of suspects.

The mix of guilty and innocent suspects would change even further, if the lineup arrays were to be as biased as those studied by researchers to date (Penrod, 2003). The bias is such that we would expect that the innocent suspect to be chosen 2–3 times as often as the average foil. If the actual multiple were 2.5, then .33 of choices from a 6-person TA lineup would be the innocent suspect (2.5/2.5 + 1 + 1 + 1 + 1 + 1). This means that the innocent suspect would be chosen by 22 witnesses from a viewing of 100 TA lineups. Consequently, the resulting pool of identifications would be 59 guilty plus 22 innocent for low-stress witnesses (73% accuracy) and 39 guilty plus 22 innocent (64% accuracy) for high stress witnesses. In short, even if heightened stress does not impact TA lineup false alarm rates, its impact on TP lineup hit rates can change materially the mix of guilty and innocent suspects identified by witnesses and the mix of correct and incorrect identifications presented to jurors.

Perhaps not surprisingly, nature of the research paradigm was also an important moderator of the effect of stress level on eyewitness identification accuracy. Mean effect size for stress level ($h = -.36$) was more than three times as great when studies were executed under the more ecologically valid conditions of the eyewitness identification paradigm than when executed within the parameters of a standard laboratory face recognition task ($h = -.10$).

Two variables were likewise identified as important moderators of the effect of heightened stress on accuracy of eyewitness recall, type of recall and witness age. Though it is not obvious why, it is clear that high stress levels impact interrogative recall much more negatively than they do narrative or free recall. Possibly the negative impact of a heightened stress level is moderated by the witness in a narrative recall situation having control over what to report and in what order. It is likewise not obvious why witness age should have emerged as a substantial moderator of the effect of heightened stress on recall, when it did not act as a moderator of stress effects on accuracy of face identification. However, there is support for the notion that measures of facial recognition and facial recall are independent, uncorrelated (Bothwell et al., 1987; Jenkins & Davies, 1985; Pigott & Brigham, 1985; Pigott, Brigham, & Bothwell, 1990). For instance, Bothwell et al. (1987) found an interaction between manipulated stress level and neuroticism on accuracy of facial identification but did not obtain a similar interaction for a measure of facial recall, description accuracy.

Our final goals for the present review were to identify any methodological or theoretical shortcomings in the body of relevant literature and to consider possible directions for future theoretical development and research. As indicated earlier, there was quite likely a major methodological difficulty with a number of studies showing apparent facilitation of memory by increases in arousal or what was defined as negative emotionality (e.g., Burke et al., 1992; Christianson, 1984; Christianson et al., 1991; Heuer & Reisberg, 1990; Libkuman et al., 1999; Safer et al., 1998). The negative emotionality manipulation in these studies generated an orienting response (arousal mode of attention control) to stimulating conditions, rather than the defensive response (activation mode of attention control) typically produced by a successful manipulation of stress or anxiety (Deffenbacher, 1994, 1999). Hence, investigators of stress effects on memory need to be concerned as to whether their experimental manipulations are eliciting the arousal mode of attention regulation or the activation mode.

There is yet another matter of methodological and theoretical importance that investigators should consider in future research, the issue of individual differences, whether they be differences in state or trait anxiety, neuroticism, specific fears, or physiological reactivity. These differences turn out to be very important. Very different patterns of response to the same stimulus situation may be shown. If an investigator were not aware of this possibility, two quite different patterns of response may cancel each other, leading to the unfortunate conclusion that an increased stress level had no demonstrable impact on memory performance. Consider an instance provided by a study (Bothwell et al., 1987) included in both of the present meta-analyses. Bothwell et al. split witnesses at the median on a scale of neuroticism, with neurotics scoring above the median and stables (emotionally

stable) scoring below. Neurotics are theorized to have very low thresholds for emotional arousal and to be predisposed to perceive a wide range of objectively non-dangerous situations as threatening and to respond with autonomic activation. Stables, on the other hand, would tend to be less anxious and physiologically reactive. The effect of stress level in their study varied dramatically with the level of neuroticism. As stress level increased from low to moderate to high, stables showed an increased level of identification accuracy, proportion correct increasing from .50 to .62 to .75, respectively. In dramatic opposition to this pattern of results was that of the neurotics; as stress level increased, proportions correct were .68, .68, and .32.

Consider one more example from another study included in both our meta-analyses, Peters' (1988) study of university students getting inoculated at a Department of Health clinic. Students were asked for physical descriptions and photo lineup identifications of both the inoculating nurse and of a second person who took their pulse 2 min later. Heart rate averaged 88 b.p.m. at inoculation versus 71 b.p.m. 2 min later. Identification accuracy overall was 66% for the second person but only 41% for the nurse. Individual differences in physiological reactivity had a profound effect on identification accuracy for the inoculating nurse. The 20 most physiologically reactive witnesses (39 b.p.m. average difference between inoculation and two minutes later) demonstrated an identification accuracy level of 31%; the 20 least physiologically reactive witnesses (3 b.p.m. difference), on the other hand, displayed an identification accuracy level of 59% for the nurse. The former witnesses clearly defined the inoculation situation as one requiring vigilance, if not actually escape or avoidance. The latter appeared to have defined the situation as more nearly one of informative perceptual intake and did not suffer the catastrophic drop in memory accuracy of the more physiologically reactive witnesses.

Hence, researchers should pay particular attention to the nature of their task. It may well be defined differently by different observers, whether it be one of simple perceptual intake or one of vigilance, for example. Whether the task be one that ordinarily produces an orienting response or a defensive response, physiological and self-report data must be examined carefully for the presence of the alternative response set in individual observers.

Thus, the modest size of the debilitating effect of heightened stress on the accuracy of eyewitness memory obtained in these meta-analyses may well be due to the averaging of its effects on two categories of witnesses, with those more anxious and physiologically reactive persons suffering a more serious drop in accuracy than those more emotionally stable persons. However, we would be remiss if we did not issue a further caveat concerning the modest obtained effect size. Whether the stress manipulation was a realistic and unexpected theft, a particularly violent video, or perhaps the threat of an injection or mild electric shock, all stress manipulations in the literature that we have examined quite likely do not reach the stress-inducing levels of extra-laboratory violent crime scenes. Thus an effect size of −.31 is perhaps a serious underestimate of the debilitating effects of stress engendered by violent crime.

For instance, results of a recent study by Ihlebaek, Love, Eilertsen, and Magnussen (2003) demonstrated that witnesses to a live staged robbery reported fewer details about the criminal event and with less accuracy than did witnesses

viewing a video recording of the same event, even though the pattern of memory errors was similar in both conditions. We agree with Ihlebaek et al. (2003) that results of laboratory studies may be an overestimate of eyewitness memory performance, especially we might add, when eyewitnesses are in a state of heightened stress.

Indeed, Morgan et al. (2004) have provided strong support for this latter caveat. They studied eyewitness capabilities of more than 500 active-duty military personnel enrolled in a survival school program. After 12 hr of confinement in a mock prisoner of war camp, participants experienced 4 hr apart, both a high-stress interrogation with real physical confrontation and a low-stress interrogation without physical confrontation; interrogations were 40 min in length. The interrogators in each instance were different individuals, with order of interrogation being counterbalanced across participants. A day after release from the prisoner of war camp, and having recovered from food and sleep deprivation, participants viewed a 15-person live lineup, a 16-person photo-spread, or a serial-presentation photo lineup of up to 16 persons. Regardless of testing method, memory accuracy for the high-stress interrogator suffered the same catastrophic decline from the level displayed for the low-stress interrogator, the same sort of catastrophic decline noted by Bothwell et al. (1987) and Peters (1988). Consider just the results from the live lineup condition: For the low stress condition, the hit and false alarm rates were .62 and .35, respectively, but comparable rates for the high stress condition were .27 and .73.

REFERENCES

References marked with an asterisk indicate studies included in the meta-analyses.

*Bailis, K. L., & Mueller, J. H. (1981). Anxiety, feedback, and self-reference in face recognition. *Motivation and Emotion, 5,* 85–96.

*Bornstein, B. H., Liebel, L. M., & Scarberry, N. C. (1998). Repeated testing in eyewitness memory: A means to improve recall of a negative emotional event. *Applied Cognitive Psychology, 12,* 119–131.

*Bothwell, R. K., Brigham, J. C., & Pigott, M. A. (1987). An exploratory study of personality differences in eyewitness memory. *Journal of Social Behavior and Personality, 2,* 335–343.

*Brigham, J. C., Maass, A., Martinez, D., & Wittenberger, G. (1983). The effect of arousal on facial recognition. *Basic and Applied Social Psychology, 4,* 279–293.

Bruck, M., Ceci, S. J., Francouer, E., & Barr, R. (1995). "I hardly cried when I got my shot!" Influencing children's reports about a visit to their pediatrician. *Child Development, 66,* 193–208.

*Buckhout, R., Alper, A., Chern, S., Silverberg, G., & Slomovits, M. (1974). Determinants of eyewitness performance on a lineup. *Bulletin of the Psychonomic Society, 4,* 191–192.

Burgwyn-Bailes, E., Baker-Ward, L., Gordon, B. N., & Ornstein, P. A. (2001). Children's memory for emergency medical treatment after one year: The impact of individual difference variables on recall and suggestibility. *Applied Cognitive Psychology, 15,* 25–48.

Burke, A., Heuer, F., & Reisberg, D. (1992). Remembering emotional events. *Memory & Cognition, 20,* 277–290.

Christianson, S.-A. (1984). The relationship between induced emotional arousal and amnesia. *Scandanavian Journal of Psychology, 25,* 147–160.

Christianson, S.-A. (1992). Emotional stress and eyewitness memory: A critical review. *Psychological Bulletin, 112,* 284–309.

Christianson, S.-A., Loftus, E. F., Hoffman, H., & Loftus, G. R. (1991). *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17,* 693–701.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

*Clifford, B. R., & Hollin, C. R. (1981). Effects of the type of incident and the number of perpetrators on eyewitness memory. *Journal of Applied Psychology, 66,* 364–370.

*Clifford, B. R., & Scott, J. (1978). Individual and situational factors in eyewitness testimony. *Journal of Applied Psychology, 63,* 352–259.

Review of the Effects of High Stress on Eyewitness Memory

*Cutler, B. L., Penrod, S. D., & Martens, T. K. (1987). The reliability of eyewitness identification: The role of system and estimator variables. *Law and Human Behavior, 11,* 233–258.

Daubert v. Merrell Dow Pharmaceuticals (1993). 509 US 579.

Deffenbacher, K. A. (1983). The influence of arousal on reliability of testimony. In S. M. A. Lloyd-Bostock & B. R. Clifford (Eds.), *Evaluating witness evidence* (pp. 235–251). Chichester, England: Wiley.

Deffenbacher, K. A. (1994). Effects of arousal on everyday memory. *Human Performance, 7,* 141–161.

Deffenbacher, K. A. (1999, July). *Negative emotionality and everyday memory: Disentangling the effects of defensive and orientation response sets.* Paper presented at the meeting of the Society for Applied Research in Memory and Cognition, Boulder, CO.

Fazey, J. A., & Hardy, L. (1988). *The inverted-U hypothesis: A catastrophe for sport psychology* (British Association for Sports Sciences Monograph No. 1). Leeds, England: National Coaching Foundation.

*Goodman, G. S., Hirschman, J. E., Hepps, D., & Rudy, L. (1991). Children's memory for stressful events. *Merrill Palmer Quarterly, 37,* 109–158.

Hardy, L., & Parfitt, G. (1991). A catastrophe model of anxiety and performance. *British Journal of Psychology, 82,* 163–178.

Hare, R. (1972). Cardiovascular components of orienting and defensive responses. *Psychophysiology, 9,* 606–614.

Hare, R., Wood, K., Britain, S., & Frazelle, J. (1971). Autonomic responses to affective visual stimulation: Sex differences. *Journal of Experimental Research in Personality, 5,* 14–22.

Hare, R., Wood, K., Britain, S., & Shadman, J. (1970). Autonomic responses to affective visual stimulation. *Psychophysiology, 7,* 408–417.

Heuer, F., & Reisberg, D. (1990). Vivid memories of emotional events: The accuracy of remembered minutiae. *Memory & Cognition, 18,* 496–506.

*Hosch, H. M., & Bothwell, R. K. (1990). Arousal, description and identification accuracy of victims and bystanders. *Journal of Social Behavior and Personality, 5,* 481–488.

Ihlebaek, C., Love, T., Eilertsen, D. E., & Magnussen, S. (2003). Memory for a staged criminal event witnessed live and on video. *Memory, 11,* 319–327.

Jenkins, F., & Davies, G. M. (1985). Contamination of facial memory through exposure to misleading composite pictures. *Journal of Applied Psychology, 70,* 164–176.

Klorman, R., Weissberg, R. P., & Wiesenfeld, A. R. (1977). Individual differences in fear and autonomic reactions to affective stimulation. *Psychophysiology, 14,* 45–51.

*Kramer, T. H., Buckhout, R., & Eugenio, P. (1990). Weapon focus, arousal, and eyewitness memory: Attention must be paid. *Law and Human Behavior, 14,* 167–184.

Lacey, J., & Lacey, B. (1974). On heart rate responses and behavior: A reply to Elliott. *Journal of Personality and Social Psychology, 30,* 1–18.

Lang, P. J., Greenwald, M. K., Bradley, M. M., & Hamm, A. O. (1993). Looking at pictures: Affective, facial, visceral, and behavioral actions. *Psychophysiology, 30,* 261–273.

Libkuman, T. M., Nichols-Whitehead, P., Griffith, J., & Thomas, R. (1999). Source of arousal and memory for detail. *Memory & Cognition, 27,* 166–190.

*Loftus, E. F., & Burns, T. E. (1982). Mental shock can produce retrograde amnesia. *Memory & Cognition, 10,* 318–323.

*Maas, A., & Kohnken, G. (1989). Eyewitness identification: Simulating the "weapon effect." *Law and Human Behavior, 13,* 397–408.

Morgan, C. A., Hazlett, G., Doran, A., Garrett, S., Hoyt, G., Thomas, P. et al. (2004). Accuracy of eyewitness memory for persons encountered during exposure to highly intense stress. *International Journal of Law and Psychiatry, 27,* 265-279.

*Mueller, J. H., Bailis, K. L., & Goldstein, A. G. (1979). Depth of processing and anxiety in facial recognition. *British Journal of Psychology, 70,* 511–515.

*Nowicki, S., Winograd, E., & Millard, B. A. (1979). Memory for faces: A social learning analysis. *Journal of Research in Personality, 13,* 460–468.

Penrod, S. D. (2003). Eyewitness identification evidence: How well are witnesses and police performing? *Criminal Justice, Spring.* 36–47, 54.

*Peters, D. P. (1988). Eyewitness memory in a natural setting. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory: Current research and issues: Vol. 1. Memory in everyday life* (pp. 89–94). Chichester, UK: Wiley.

*Peters, D. P. (1991). The influence of stress and arousal on the child witness. In J. Doris (Ed.), *The suggestibility of children's recollections* (pp. 60–76). Washington, DC: American Psychological Association.

*Peters, D. P. (1997). Stress, arousal, and children's eyewitness memory. In N. L. Stein, P. A. Ornstein, B. Tversky, & C. J. Brainerd (Eds.), *Memory for everyday and emotional events.* Mahwah, NJ: Erlbaum.

Peterson, C., & Bell, M. (1996). Children's memory for traumatic injury. *Child Development, 67,* 3045–3070.

Peterson, C., & Whalen, N. (2001). Five years later: Children's memory for medical emergencies. *Applied Cognitive Psychology, 15,* 7–24.

Pigott, M. A., & Brigham, J. C. (1985). Relationship between accuracy of prior description and facial recognition. *Journal of Applied Psychology, 70,* 547–555.

Pigott, M. A., Brigham, J. C., & Bothwell, R. K. (1990). A field study on the relationship between quality of eyewitnesses' descriptions and identification accuracy. *Journal of Police Science and Administration, 17,* 84–88.

Quas, J. A., Goodman, G. S., Bidrose, S., Pipe, M., Craw, S., & Ablin, D. S. (1999). Emotion and memory: Children's long-term remembering, forgetting, and suggestibility. *Journal of Experimental Child Psychology, 72,* 235–270.

*Read, J. D., Yuille, J. C., & Tollestrup, P. (1992). Recollections of a robbery: Effects of arousal and alcohol upon recall and person identification. *Law and Human Behavior, 16,* 425–446.

Rosenthal, R. (1995). Writing meta-analytic reviews. *Psychological Bulletin, 118,* 183–192.

Safer, M. A., Christianson, S.-A., Autry, M. W., & Osterlund, K. (1998). Tunnel memory for traumatic events. *Applied Cognitive Psychology, 12,* 99–117.

Shapiro, P. N., & Penrod, S. D. (1986). A meta-analysis of facial identification studies. *Psychological Bulletin, 100,* 139–156.

*Shrimpton, S., Oates, K., & Hayes, S. (1998). Children's memory of events: Effects of stress, age, time delay and location of interview. *Applied Cognitive Psychology, 12,* 133–143.

*Siegel, J. M., & Loftus, E. F. (1978). Impact of anxiety and life stress upon eyewitness testimony. *Bulletin of the Psychonomic Society, 4,* 191–192.

*Stanny, C. J., & Johnson, T. C. (2000). Effects of stress induced by a simulated shooting on recall by police and citizen witnesses. *American Journal of Psychology, 113,* 359–386.

Steblay, N. M. (1997). Social influence in eyewitness recall: A meta-analytic review of lineup instruction effects. *Law and Human Behavior, 21,* 283–297.

*Tooley, V., Brigham, J. C., Maas, A., & Bothwell, R. K. (1987). Facial recognition: Weapon effect and attentional focus. *Journal of Applied Social Psychology, 17,* 845–859.

Tucker, D. M., & Williamson, P. A. (1984). Asymmetric neural control systems in human self-regulation. *Psychological Review, 91,* 185–215.

*Zanni, G. R., & Offermann, J. T. (1978). Eyewitness testimony: An exploration of question wording upon recall as a function of neuroticism. *Perceptual and Motor Skills, 46,* 163–166.

Law and Human Behavior
June, 2011

Original Article

*200 AFTER 30 YEARS, WHAT DO WE KNOW ABOUT WHAT JURORS KNOW? A META-ANALYTIC REVIEW OF LAY KNOWLEDGE REGARDING EYEWITNESS FACTORS

Sarah L. **Desmarais** [FNa1]

J. Don Read [FNa2]

Abstract Surveys typically characterize lay knowledge of eyewitness factors as low and highly variable. However, there are notable differences across methodologies, samples, and individual factors. To examine these differences systematically, we took a meta-analytic approach to reviewing the findings of 23 surveys assessing lay knowledge of eyewitness issues. Our analyses examined the beliefs of 4,669 respondents. Overall, respondents correctly agreed with survey items approximately two-thirds of the time. Results revealed significant differences in performance as a function of variable type, question format, and over time. We found few differences as a function of sample type, publication status, or jurisdiction. Although performance varied, a majority of lay respondents achieved "correct" consensus for as many as 11 of the 16 items included in this review.

Although the exact criteria vary, juror knowledge is always central to the decision of whether expert opinion on the topic of eyewitness testimony should be admissible in court. Specifically, the judge must ask him or herself whether the expert witness testimony would assist triers of fact in making a decision of acceptable reliability. When faced with this question, judges regularly rule that many components of proffered eyewitness expert testimony are common sense and known to jurors (Benton, McDonnell, Ross, Thomas, & Bradshaw, 2007). In contrast, most surveys report that student and community respondents demonstrate frequent disagreement with both the opinions of experts and the outcomes of empirical research (e.g., Benton, Ross, Bradshaw, Thomas, & Bradshaw, 2006; Deffenbacher & Loftus, 1982; Kassin & Barndollar, 1992; Schmechel, O'Toole, Easterly, & Loftus, 2006; Yarmey & Jones, 1983; but see Read & **Desmarais**, 2009a). As a result, authors have concluded almost unanimously that juror knowledge is an inadequate basis for understanding the facts of a case involving eyewitness testimony. For instance, finding significant differences between juror and expert responding on 26 out of 30 items, Benton et al. (2006) asserted that "the discrepancy between lay understanding of factors affecting eyewitness accuracy and what decades of empirical research has reliably demonstrated to be true continues to be evidenced" and that "jurors ... exhibit important limitations in their knowledge of eyewitness issues, their knowledge diverges significantly from expert opinion, and it is not high in overall accuracy" (p. 126).

Such findings have led many researchers and legal scholars to argue that expert testimony on eyewitness factors should not be ruled inadmissible as a matter of course (Benton et al., 2006, 2007; Cutler & Penrod, 1995; Schmechel et al., 2006). However, closer examination of the findings reveals notable differences across methodologies, samples, and individual eyewitness issues. Take, for example, results of the Benton et al. (2006) study. Although it is true that, overall, there was significant disagreement between lay and expert responses, there also were a number of items upon which agreement rates did not differ

significantly, such as alcohol intoxication *201 (experts: 90% vs. jurors: 96%), stress (experts: 60% vs. jurors: 68%), trained observers (experts: 39% vs. jurors: 28%), and event violence (experts: 36% vs. jurors: 26%). If we compare Benton et al.'s lay responses to what 80% of experts consider "correct," there are another three items upon which jurors appear to be well informed. Specifically, lay agreement rates exceeded 80% for the wording of questions, child suggestibility, and attitudes & expectations items. Although the consensus criterion of 80% has been adopted for establishing general agreement among experts (Kassin, Ellsworth, & Smith, 1989; Kassin, Tubb, Hosch, & Memon, 2001), we are hesitant to adopt a similar criterion for general agreement amongst jurors because they answered a different kind of question and selected among different response options than experts.

The more recent work of Read and **Desmarais** (2009a) questioned the validity of comparing lay responses to those of experts as the basis for conclusions regarding the necessity of eyewitness testimony. Considering responses of three large community surveys (total $N = 999$), the authors concluded that lay responses frequently approximated those of experts. Correct agreement rates averaged 70% across these three surveys, a marked improvement over the rates observed in past studies (e.g., Benton et al., 2006: 51%; Kassin & Barndollar, 1992: 61%; Yarmey & Jones, 1983: 30%). Rather than statistically comparing lay judgments of each statement's validity to the Kassin et al. (2001) experts' responses regarding whether the topic/proposition had "sufficient reliability" to be presented in court, Read and **Desmarais** used experts' judgments of the reliability of research evidence in support of each empirical proposition. Examining the 18 propositions that received 80% consensus in reliability judgments from experts, they found 12 topics that received "correct" agreement among lay respondents at rates of 70% or greater. There were another four items upon which both experts and lay participants agreed at rates between 60 and 70%. Nonetheless, the authors noted significant deficiencies in knowledge for 50% of the topics assessed.

Given the centrality of juror knowledge to the admissibility of expert eyewitness testimony, there is a need to clarify which topics are "beyond the ken" of a jury. Although researchers have surveyed lay beliefs regarding eyewitness issues since the late 1970s, there has been no systematic review of the findings. This research represents an attempt to do just that. Before we proceed, however, a conceptual difference must be drawn between jurors' knowledge and their use of such knowledge. Over the past 30 years, researchers have used both direct and indirect methods to assess lay knowledge of eyewitness issues. Direct methods emphasize a survey approach in which respondents answer questions about the general effects of eyewitness variables, whereas indirect methods examine the ways in which lay participants make use of their knowledge about eyewitness factors. Depending upon their subsequent evaluations of eyewitness credibility and, ultimately, the verdict rendered, researchers then draw inferences regarding jurors' pre-trial beliefs about relevant eyewitness factors (e.g., Cutler, Penrod, & Dexter, 1989; Martire & Kemp, 2009). With direct methods, we answer the question: What do jurors know about eyewitness issues? Indirect methods, in contrast, ask: How do jurors use their knowledge of eyewitness issues? Because most studies speak to the former question and not the latter, we focus our research on direct methods.

Focusing on surveys to the exclusion of other research approaches, we nonetheless face the challenging task of making sense of findings from over two dozen studies. Every decade or so, there has been a peak in interest for (re-) surveying lay knowledge. The first wave of surveys emerged in the late 1970s and early 1980s (e.g., Deffenbacher & Loftus, 1982; Loftus, 1979; Yarmey & Jones, 1983). These surveys typically used multiple-choice items to assess lay beliefs and directly compared these responses to those of experts on identical items. The first of its kind, the Knowledge of Eyewitness Behavior Questionnaire (KEBQ) is a 14-item, four-alternative, multiple-choice questionnaire designed to assess knowledge of variables that influence eyewitness memory. Deffenbacher and Loftus (1982) administered the KEBQ to two samples of college students and two samples of jury-eligible citizens in Washington, D.C. Across samples, Deffenbacher and Loftus (1982) found that the majority of respondents did not give the correct answer to a number of the items. Administering the same questionnaire to survey British and Australian samples, Noon and Hollin (1987) and McCon-key and Roche (1989) arrived at similar conclusions. Noon and Hollin, however, did note significant differences between British and American respondents at the item level suggesting that knowledge may differ as a function of jurisdiction or over time. Using a similar 16-item multiple-choice questionnaire, Yarmey and Jones' (1983) surveys of Canadian respondents were no more favorable. Again, the authors concluded that knowledge of variables that influence eyewitness accuracy is *not* common sense.

Other researchers used agree-disagree items to survey lay knowledge of eyewitness issues. In the late 1980s, Kassin et al. (1989) published results of an eyewitness expert survey focused on establishing which topics meet the 'general acceptance'

standard of the *Frye* test. Kassin and Barndollar (1992) subsequently used the 21 statements from this expert survey to assess student and community respondents' beliefs regarding the relationship between eyewitness variables and memory performance. As in earlier surveys, the researchers found that a significant proportion of participants did not agree with items that *202 were accurate. Furthermore, although there are good reasons for anticipating differences across samples (e.g., age, familiarity with test-taking procedures, *cf.* Bornstein, 1999), student and community responses were highly similar. This finding suggests that students and nonstudent adults are not differentially sensitive to factors that influence eyewitness accuracy (Kassin & Barndollar, 1992). Kassin et al. (2001) re-assessment of expert beliefs included 17 of the original 21 items as well as 13 new ones. Again, there was quick uptake of the items by researchers interested in assessing lay knowledge and comparing lay and expert opinion (e.g., Benton et al., 2006; Read & **Desmarais**, 2009a). As mentioned earlier, studies using the Kassin et al. (2001) items have shown strikingly different results and provided the impetus to thoroughly review the survey literature on the topic.

The Present Research

After 30 years, what do we know about what jurors know? To answer this question, we used a meta-analytic approach to summarize the results of 23 surveys assessing lay beliefs of eyewitness issues. We specifically examined responses as function of eyewitness variable type, sample type, question format, jurisdiction, and year of administration to test possible explanations for inconsistent findings across studies. We focus our analyses on the 16 eyewitness topics for which the Kassin et al. (2001) experts achieved 80% consensus regarding whether the phenomena are reliable enough for psychologists to present in courtroom testimony. There are persuasive arguments for using a different criterion, and thus, a somewhat different subset of topics (see Read & **Desmarais**, 2009a). However, these 16 are most frequently included in past research and have demonstrated the greatest consistency across experts' responses to different questions about them. Furthermore, without an 80% level of agreement between experts, courts may be unlikely to admit testimony on the topic. If not admitted, the results of surveys regarding lay knowledge would not likely be heard in court.

Method

Selection Criteria

We located studies through a number of sources: (a) searching the PsycINFO database; (b) reviewing programs for posters and papers presented at relevant scholarly meetings (e.g., American Psychology-Law Society, Society for Applied Research in Memory and Cognition); and (c) contacting researchers in the field who may have knowledge of unpublished literature. We included only those studies that used direct methods to survey lay knowledge. We excluded studies that presented results obtained using indirect methods, presented data in aggregate only (i.e., collapsing across items), or used a question format other than agree-disagree or multiple-choice. For instance, Lindsay (1994) asked participants to rate the likelihood that an eyewitness would make an accurate identification under a series of varying circumstances, and Shaw, Garcia, & McClure (1999) asked participants to indicate factors they believe affect the accuracy of eyewitness testimony in a free-response format. Although the findings of these two studies certainly speak to lay beliefs regarding eyewitness issues, they are not directly comparable to those of other studies that have assessed performance on agree-disagree and multiple-choice items. Within each study, we only included data attributable to "lay" respondents, excluding responses of experts or legal professionals (e.g., judges, lawyers, law enforcement personnel).

Sample of Surveys

In total, we identified 23 surveys from 15 studies, representing the responses of 4,669 lay respondents. Of those, approximately two-thirds (67%, $N = 3,111$, $k = 12$) were community volunteers and one-third (33%, $N = 1,479$, $k = 10$), university students. One study (Kassin & Barndollar, 1992) surveyed both without distinguishing between community and student respondent data. Sixteen surveys comprised multiple-choice items, six comprised agree-disagree items, and one (Rahaim & Brodsky, 1982) included both. Eleven surveys were conducted in the United States, representing approximately half of respondents ($N = 2,411$, 52%). A significant minority of respondents were Canadian: seven surveys, representing 1,838 re-

spondents (39%). The remaining respondents were surveyed in the United Kingdom (three surveys, 5% of respondents, $N =$ 249) and Australia (two surveys, 4% of respondents, $N = 171$).

Outline of Analyses

We calculated average item performance with respect to samples (unweighted means) and respondents (means weighted by survey sample size). We also calculated performance across items at the survey level. We report lay performance overall and as a function of variable type (system, estimator), sample type (community, student), question format (multiple-choice, agree-disagree), jurisdiction (the United States, Canada, the United Kingdom, and Australia), and year of administration (collapsed for presentation into the late 1970s/early 1980s, $k = 7$; late *203 1980s/early 1990s, $k = 6$; and late 1990s/early 2000s, $k = 10$). In addition to these methodological and theoretical variables of interest, we examined publication status (published, unpublished) and associations between the variables of interest that may confound our comparisons.

Results

We present mean rates of correct responding by item in Table 1. For a summary of item-level responding within surveys, see Read and **Desmarais** (2009b, Tables 6.1 and 6.2). As may be seen in Table 1, individual surveys generally included only a selection of the eyewitness factors. The modal number of surveys assessing any given topic was nine (range = 2-30). Only one topic was assessed in all 23 surveys: cross-race bias. Three other topics were assessed in greater than two-thirds of the surveys: mugshot bias ($N = 2,267$, $k = 17$), question wording ($N = 3,376$, $k = 19$), and accuracy-confidence ($N = 4,123$, $k = 21$). The vast majority of participants (>=70%) also responded to items pertaining to unconscious transference ($N = 3,386$, $k = 11$) and confidence malleability ($N = 3,110$ $k = 9$). The average number of respondents per item was 2,679.69 ($SD = 964.04$; range = 564-4,669).

Table 1 Mean rates of percent correct agreement with Kassin et al. (2001) experts

Eyewitness topics	Kassin et al. (2001) experts		Lay respondents			
	Research evidence is reliable[FNa]	Reliable enough for courtroom testimony[FNb]	k	N	Unweighted means	Weighted means[FNc]
System variables	92	93			67	71
Confidence malleability	97	95	9	3,110	73	81
Lineup instructions	90	98	9	2,182	70	75

Mug-shot-induced bias	98	95	17	2,267	62	65
Presentation format	74	81	4	564	46	49
Question wording	100	98	19	3,376	85	87
Estimator variables	89	88			64	67
Accuracy & confidence	90	87	21	4,123	43	51
Alcohol intoxication	83	90	9	2,300	91	92
Attitudes & expectation	100	92	9	2,182	85	88
Child sug-gestibility	94	94	9	2,300	76	75
Cross-race bias	95	90	23	4,669	50	57
Exposure time	83	81	10	2,379	61	65
Forgetting curve	79	83	8	2,130	58	61

Hypnotic suggestibility	84	91	8	2,130	51	53
Post event information	95	94	10	2,379	72	74
Unconscious transference	92	81	11	3,386	63	69
Weapon focus	89	87	19	3,398	53	52
Overall	90	90			65	68

FNa. Calculated by summing expert responses of "tends to favor", "generally reliable", and "very reliable" regarding how they would characterize the reliability of the phenomena (see Kassin et al., 2001, Table 3)
FNb. Expert judgments regarding whether the phenomenon was sufficiently reliable to be presented in court (see Kassin et al., 2001, Table 4)
FNc. Overall means weighted as a function of survey sample size

Review of Table 1 reveals significant variation in lay performance across items. On the one hand, there are some items for which lay knowledge approximates expert responding, such as alcohol intoxication, attitudes & expectations, and question wording. On the other hand, lay knowledge is clearly deficient for other topics, such as presentation format, the accuracy-confidence relationship, and hypnotic suggestibility. The unweighted mean correct response rate was 65.0% ($SD = 14.70$), ranging from 43% for accuracy-confidence to 91% for alcohol intoxication. The weighted mean was slighter higher at 68.4% ($SD = 14.04$), ranging from 46% for presentation format to 92% for alcohol intoxication.

We present mean rates of correct responding within surveys in Table 2. There was a wide range of correct agreement rates across surveys, from a low of 24.0% for Rahaim and Brodsky (1982) to a high of 80.0% for Read and **Desmarais** (2009a, Survey 3). The mean within-survey correct response rate was 57.8% ($SD = 14.88$).

Table 2 Summary of surveys included

Survey	Sample type	N	Question format	Mean % correct responding		
				Overall	System variables	Estimator variables

Australia						
McConkey and Roche (1989) Sample 1	U		124	MC	49.7	69.0	35.3
McConkey and Roche (1989) Sample 2	U		47	MC	73.3	78.7	69.3
Canada						
Desmarais and Read (2008) Sample 1	U		270	MC	77.2	87.3	73.6
Desmarais and Read (2008) Sample 2	C		449	MC	64.9	72.5	62.2
Read and **Desmarais** (2009a) Survey 1	C		201	AG	73.5	77.2	71.8
Read and **Desmarais** (2009a) Survey 2	C		200	AG	75.8	82.6	72.6
Read and **Desmarais** (2009a) Survey 3	C		598	AG	80.0	87.0	78.1
Yarmey and Jones (1983)			60	MC	52.0	61.5	45.7

35 LHUMB 200
35 Law & Hum. Behav. 200

Yarmey and Jones (1983)	C	60	MC	47.8	50.0	46.3
United Kingdom						
Hope et al. (2009)	C	197	MC	62.8	64.5	62.4
Noon and Hollin (1987) Sample 1	U	28	MC	54.0	73.7	39.3
Noon and Hollin (1987) Sample 2	C	24	MC	49.7	60.0	42.0
United States						
Benton et al. (2006)	C	111	AG	52.7	53.2	52.5
Deffen- bacher and Loftus (1982) Sample 1	U	76	MC	54.3	73.3	40.0
Deffen- bacher and Loftus (1982) Sample 2	U	100	MC	47.7	63.0	36.3
Deffen- bacher and Loftus (1982) Sample 3	C	46	MC	37.0	54.0	20.0

35 LHUMB 200
35 Law & Hum. Behav. 200

Devenport and Cutler (2009)	U	222	MC	74.9	74.3	75.1
Kassin and Barndollar (1992)	C; U	79	AG	58.3	79.0	54.2
Lane et al. (2008)	U	52	MC	51.1	51.2	51.1
Loftus (1979)	U	500	MC	61.3	90.0	47.0
Rahaim and Brodsky (1982)	C	28	AG; MC	24.0	-	24.0
Schmechel et al. (2006)	C	1007	AG	73.0	85.5	64.7
Seltzer et al. (1990)	C	190	MC	34.5	-	34.5

C community, *U* university students. *MC* multiple-choice, *AG* agree-disagree (or some similar variation). Means calculated based on the 16 items included in this meta-analytic review

*204 Interrelation Between Variables of Interest

We conducted a series of analyses to test for associations between the variables of interest, as well as sample size. These analyses also afforded the opportunity to examine trends in research approaches and methodologies. As may be seen in Table 3, there are some significant associations that we will need to consider in subsequent analyses. In particular, the number of system variables included in each survey increased with the number of estimator variables (indicating an overall increase in survey length) and over time. Sample size also increased significantly over time. The number of estimator variables included was

higher for agree-disagree ($M = 9.50$, $SD = 3.21$) compared to multiple-choice surveys ($M = 5.12$, $SD = 3.44$), $t(21) = 2.72$, $p < .05$, $d = 1.19$, 95% CI [1.0, 7.7], as well as for published ($M = 5.53$, $SD = 3.81$) compared to unpublished surveys ($M = 9.75$, $SD = 1.50$), $t(21) = 2.15$, $p < .05$, $d = 0.94$, 95% CI [0.1, 8.3]. A Mann-Whitney U test revealed that sample size was significantly higher for agree-disagree (M Rank $= 16.67$) than multiple-choice surveys (M rank $= 10.35$), $z = 1.96$, $p = .05$. Researchers used agree-disagree formats much more frequently for community samples, $^2(1, N = 22) = 5.39$, $p < .05$, $= .50$. Indeed, all surveys of university students comprised multiple-choice items compared to only 58.3% of community surveys. There also have been differing trends in question format over time, with a peak in use of agree-disagree items in the late 1980s/early 1990s: 83.3% compared to none in the late 1970s/early 1980s and 50.0% in the late 1990s/early 2000s, $(2, N = 23) = 5.71$, $p = .06$, $= .50$. There were no systematic associations between jurisdiction and sample size, $F(3, 19) = 0.56$, $p = .65$, $>>^2_p = .08$, number of system or estimator *205 variables, $F\text{'s}(3, 19) <= 2.26$, $p\text{'s} >= .24$, $^2_p\text{'s} <= .20$, or the other independent variables, $^2\text{'s}(3, N\text{'s} = 22\text{-}23) <= 3.47$, $p\text{'s} >= .33$, $\text{'s} <= .40$. [FN1]

Table 3 Interrelation between variables of interest

Variables	Correlations						
	1	2	3	4	5	6	7
1. Number of system variables	-						
2. Number of estimator variables	.77***	-					
3. Sample type	-.15	.12	-				
4. Question format	-.34	-.51*	-.50*	-			
5. Year of administration	.65***	go***	.25	-.49*	-		
6. Publication status	-.29	-.43*	.04	-.27	-.47*	-	

7. Sample size	-.04	.13	.24	-.42*	.41*	-.08	-

Sample type: student = 0; community = 1. Question format: agree-disagree = 0; multiple-choice = 1. Year of administration: 0 = late 1970s/early 1980s; 1 = late 1980s/early 1990s; 2 = late 1990s/early 2000 s. Publication status: 0 = unpublished; 1 = published. * $p < .05$. ** $p < .01$. *** $p < .001$

Variable Type

Overall, lay performance was consistently better for system than estimator variables (see Tables 1 and 2). Although highly correlated ($r = .79, p < .001$), a repeated measures ANOVA revealed that mean performance was significantly better for system ($M = 70.8\%, SD = 12.59$) compared to estimator variables ($M = 54.3\%, SD = 16.00$), $F(1, 20) = 31.94, p < .001, \eta^2_p = .62$. Controlling for the number of system and estimator variables included in each survey contributed to a slight increase in effect size ($\eta^2_p = .65$).

Sample Type

To compare performance between student and community samples, we conducted pairwise comparisons. We excluded the Kassin and Barndollar (1992) results from these analyses because, as noted earlier, the authors did not differentiate between student and community responses. Review of Table 4 reveals that student and community responding were highly similar. In fact, we found no significant differences overall or for system and estimator variables, $t's(18-20) <= 0.61, p's >= .55, d's <= 0.27$, 95% CIs [-10.6, 17.1], [-8.7, 15.7], and [-17.2, 14.5], in order. Differences remained non-significant after controlling for the number of system and estimator variables surveyed, $F's(1, 17-19) <= 0.95, p's >= .34, \eta^2_p's <= .05$.

[Note: The following table/form is too wide to be printed on a single page. For meaningful review of its contents the table must be assembled with part numbers in ascending order from left to right. Row numbers, which are not part of the original data, have been added in the margins and can be used to align rows across the parts.]

****
************** This is piece: 1
****

1	**Table 4** Mean percent correct agreement with survey items by sample type, question format, jurisdiction, and year of administration								
2	Eye-witness topics	Sample type		Question format		Jurisdiction				Year of administration
3		Stu-dent	Com munity	Agree-disagee	Multi-ple-choice	Aus-tralia	Can-ada	Unite d King-dom	Unite d States	Late early

4	System variables	64.4	67.9	77.4	68.2	73.8	74.0	66.1	69.3	65.3
5	Confidence malleability	63.3	76.9	76.9	66.8	-	83.0	-	59.1	-
6	Lineup instructions	67.7	71.4	70.6	68.8	-	82.2	-	54.3	-
7	Mug shot-induced bias	62.5	63.0	79.3	59.2	64.3	64.7	60.7	61.4	52.1
8	Presentation format	40.0	47.3	47.3	40.0	-	55.5	35.5	45.5	-
9	Question wording	88.1	80.6	90.4	83.1	93.0	85.3	76.3	86.3	79.4
10	Estimator variables	58.5	60.9	85.6	47.3	52.3	64.3	47.4	45.4	37.0
11	Accuracy & confidence	41.2	44.5	57.8	37.5	41.5	57.4	22.3	39.7	30.2
12	Alcohol intoxication	89.3	91.2	90.5	90.6	-	91.0	92.0	89.3	-

13	Attitudes & expectation	84.0	85.2	86.4	83.8	-	88.6	-	81.0	-
14	Child suggestibility	73.7	76.8	83.5	69.6	-	78.0	64.0	76.0	-
15	Cross-race bias	54.5	50.6	61.0	48.6	63.0	63.3	51.0	42.7	38.0
16	Exposure time	58.3	66.3	60.6	61.4	-	67.6	68.0	51.0	-
17	Forgetting curve	76.5	54.0	49.2	72.7	-	61.0	-	53.0	-
18	Hypnotic suggestibility	59.5	48.2	48.4	54.7	-	55.2	-	43.3	-
19	Post event information	71.3	72.3	78.0	66.6	-	79.6	60.0	66.3	-
20	Unconscious transference	52.7	67.7	66.8	59.2	-	67.6	77.0	56.4	-
21	Wea	49.5	56.9	59.6	50.9	50.0	61.3	49.7	47.4	50.0

		1970s/1980s	Late early	1980s/1990s	Late early	1990s/2000s			
	pon focus								
22	Over all	63.5	65.8	68.9	53.9	61.5	67.3	55.5	51.7	46.3
23	Mean percent correct agreement overall, and system and estimator variables calculated based on item means. -, Item not included in surveys								

****

************** This is piece: 2

****

	1970s /1980s	Late early	1980s/ 1990s	Late early	1990s/ 2000s
1					
2					
3					
4	72.1		73.5		
5	-		72.4		
6	68.0		70.0		
7	61.9		69.9		
8	-		45.5		
9	87.8		86.7		
10	45.7		66.4		
11	33.8		55.5		

12	-	90.6
13	89.0	84.8
14	-	75.8
15	48.8	63.3
16	37.0	63.7
17	41.0	60.4
18	46.0	51.4
19	75.0	72.0
20	65.0	63.2
21	52.4	55.3
22	53.3	68.6
23		

Patterns of responding to individual items were mixed, but the differences, again, were non-significant, t's(18-20) $<= 0.61$, p's $>= .12$, d s $<= 0.30$ (see Table 4). University students outperformed community samples on four items: question wording, cross-race bias, forgetting curve, and hypnotic suggestibility. Community samples outperformed university students on 10 items: confidence malleability, lineup instructions, presentation format, accuracy-confidence, alcohol intoxication, attitudes & expectations, child suggestibility, exposure time, unconscious transference, and weapon focus. Student and community mean agreement rates were nearly identical for mugshot bias and post event information.

Question Format

Few surveys have examined the impact of question format on measurement of lay knowledge (for exceptions see Hope et al., 2009, and Read & **Desmarais**, 2009a). However, the question format literature provides good evidence that we should anticipate differential accuracy associated with the various response formats because of differences in chance performance rates; for example, better performance on agree-disagree compared to multiple-choice items. Overall, our results support this hypothesis (see Table 4). Pairwise comparisons demonstrated significantly better performance on agree-disagree surveys overall, $t(21) = 2.32, p < .05, d = 1.01, 95\%$ CI [1.6, 28.4], and on estimator variables in particular, $t(21) = 2.55, p < .05, d = 1.11$, 95% CI [3.4, 33.3]. We found the same pattern for system variables; however, the difference did not reach statistical significance, $t(19) = 1.57, p = .13, d = 0.75, 95\%$ CI [-3.1, 21.5]. Controlling for the number of estimator variables surveyed and year of administration, the differences between question formats for overall survey and estimator means were no longer significant, F's(1, 20) <= 0.80, p's >= .38, $^2{}_p$'s <= .04. The difference between *206 multiple-choice and agree-disagree formats surveys remained non-significant for system variables after controlling for year of administration, $F(1, 18) = 1.17, p = .29, ^2{}_p = .06$.

As may be seen in Table 4, item-level comparisons were significant for three of the 16 factors. Respondents performed better on the agree-disagree compared to multiple-choice format for mugshot bias, accuracy-confidence, and child suggestibility, t's(7-19) >= 2.20, p's < .05, d's >= 1.46, 95% CIs [5.0, 35.3], [1.0, 39.7], and [5.7, 22.1], in order. In contrast, meanrates of correct responding to forgetting curve items were lower for the agree-disagree format, $t(6) = 2.39, p = .05, d = 1.95, 95\%$ CI [-47.5, 0.5]. Hypnotic suggestibility was the only other item for which we saw this reverse pattern, but the difference was not significant, $t(6) = 0.67, p = .53, d = 0.53, 95\%$ CI [-29.2, 16.7].

Jurisdiction

To examine responding as a function of jurisdiction, we conducted one-way ANOVAs with jurisdiction as the grouping variable. There were very few significant differences overall (see Table 4). [FN2] Specifically, the overall study means, system variable means, and estimator variable means did not differ significantly, F's(3, 17-19) <= 2.14, p's >= .13, $^2{}_p$'s <= .25. Performance of Canadian compared to American respondents was significantly better for four items: confidence malleability, lineup instructions, presentation format, and cross-race bias, t's(2-16) >= 2.65, p's < .05, d's >= 1.45, 95% CIs [4.6, 43.2], [3.0, 52.9], [0.5, 39.5], and [5.5, 35.6], in order. One other significant difference emerged: for accuracy-confidence, Canadian respondents outperformed British respondents on average, $t(8) = 2.79, p < .05, d = 1.97, 95\%$ CI [6.0, 64.1].

Year of Administration

To examine whether lay knowledge of eyewitness issues has changed over time, we conducted one-way ANOVAs with year of administration as the grouping variable. Results revealed significant increases in lay performance over time, $F(2, 20) = 8.33, p < .01, ^2{}_p = .45$. As may be seen in Table 4, overall survey means increased significantly between the late 1970s/early 1980s and late 1990s/early 2000s, and between the late 1980s/early 1990s and late 1990s/early 2000s, t's(14-5) >= 2.66, p's < .05, d's >= 1.33, 95% CIs [-34.0, -10.6] and [-27.2, -3.0], respectively. Controlling for the number of system and estimator variables, the main effect of time on overall *207 survey means was no longer significant, $F(4, 18) = 0.97, p = .40, ^2{}_p = .10$. In contrast, effects remained significant after controlling for presentation format and publication status, with only small changes in effect size (to $^2{}_p = .34$ and $^2{}_p = .46$, respectively). There were no significant improvements in performance between the late 1970s/early 1980s and late 1980s/early 1990s, $t(11) = 1.00, p = .34, d = 0.55, 95\%$ CI [-22.2, 8.3].

A repeated measures ANOVA with system and estimator means as the within-survey variables revealed a significant effect of time, $F(2, 18) = 5.40, p < .05, >>^2{}_p = .38$, that was qualified by a time x variable type interaction, $F(2, 18) = 8.16, p < .01, ^2{}_p = .48$ (see Table 4). Like the overall survey means, estimator variable means increased significantly between the late 1970s/early 1980s and late 1990s/early 2000s, and between the late 1980s/early 1990s and late 1990s/early 2000s, t's(14-15) >= 3.62, p's < .01, d's >= 1.42, 95% CIs [-40.0, -18.8] and [-32.9, -8.4], respectively. Controlling for presentation format and publication status resulted in slight decreases in effect size (to $^2{}_p = .46$ and $^2{}_p = .39$, respectively). There were no significant improvements in mean performance on estimator variables between the late 1970s/early 1980s and late 1980s/early 1990s,

$t(11) = 1.28$, $p = .23$, $d = 0.70$, 95% CI [-23.7, 6.3]. Although there was a trend for average performance on system variables to increase over time, the main effect was not significant, $F(2, 18) = 0.81$, $p = .46$, $\eta^2_p = .08$, and remained so after controlling for presentation format and publication status, $F's(3, 17) <= 1.20$, $p's >= .33$, $\eta^2_p 's <= .12$.

At the item level, lay performance improved significantly over time for three eyewitness phenomena: mugshot bias, the accuracy-confidence relationship, and cross-race bias, $F's(2, 14-20) >= 3.61$, $p <= .05$, $\eta^2_p 's >= .33$ (see Table 4). Although there was a trend for lay performance to improve between the late 1970s/early 1980s and late 1980s/early 1990s, the differences were not significant, $t's(7-11) <= 1.81$, $p's >= .11$, $d's <= 1.23$, 95% CIs [-22.6, 3.0], [-26.1, 18.9], and [-28.2, 6.5], in order. We found the same pattern of results between surveys conducted in the late 1980s/early 1990s and late 1990s/early 2000s, with the difference approaching significance for accuracy-confidence and cross-race bias, $t(14)'s >= 2.07$, $p's = .06$, $d's >= 1.11$, 95% CIs [-44.1, 0.7] and [-29.3, 0.4]. For all three items, improvements in performance from the late 1970s/early 1980s to the late 1990s/early 2000s were significant, $t's(11-15) >= 2.47$, $p's < .05$, $d's >= 1.49$, 95% CIs [-33.6, - 1.9], [-44.9, -5.7], and [-38.0, -12.6].

Publication Status

Our final set of analyses examined whether there were any systematic differences as a function of publication status. We found only one difference between published and unpublished surveys: Mean rates of correct agreement were significantly higher on the child suggestibility item for published compared to unpublished surveys (80.8% vs. 69.5%), $t(7) = 2.45$, $p < .05$, $d = 1.85$, 95% CI [-22.2, -0.4].

Discussion

The goal of this research was to make sense of the findings from 30 years of surveys examining lay knowledge of eyewitness issues. To do so, we conducted a meta-analytic review of 23 surveys assessing lay beliefs of eyewitness issues with regard to a number of methodologically and theoretically relevant variables. We focused our analyses on the 16 eyewitness topics for which the Kassin et al. (2001) experts achieved 80% consensus regarding whether the phenomena are sufficiently reliable for psychologists to present in courtroom testimony. Overall, respondents correctly agreed with survey items approximately two-thirds of the time. Results demonstrated significant differences in performance as a function of variable type and question format, and identified significant changes over time. There were few differences as a function of sample type, publication status, or jurisdiction. We discuss these findings in more detail below.

Across surveys, respondents demonstrated lesser knowledge regarding factors that are not under the control of the criminal justice system and for which their impact on witness testimony can only be estimated (i.e., estimator variables) compared to factors that are under the control of the criminal justice system (i.e., system variables) (cf. Wells, 1978). As recently noted by Benton et al. (2006), this finding is inconsistent with research demonstrating a bias for lay explanations of eyewitness phenomena to focus on estimator issues (Boyce, Beaudry, & Lindsay, 2007). This pattern of results is consistent, however, with Benton et al.'s results across a larger number of variables (eight system and 22 estimator variables) and with the focus on system variables in eyewitness identification research (cf. Wells & Olson, 2003). System variables may be easier to understand because the means through which their influence can be managed are tangible and concrete, such as altering lineup procedures. Understanding the processes through which estimator variables influence eyewitness accuracy is undoubtedly a more difficult task, given the abstract nature of many of these factors as well as the variability of their impact in specific cases (Wells & Olson, 2003). Regardless the reason for the poor performance, estimator variables more frequently appear to be "beyond the ken" of a jury.

*208 The differential responding across question format emphasizes the importance of using multi-method approaches to examining lay knowledge. This finding also raises interesting questions about some of the fundamental assumptions in this area of work. Do these surveys accurately reflect what lay respondents believe or know about eyewitness phenomena? The improved performance on agree-disagree compared to multiple-choice items may be attributable to the increased likelihood of guessing the correct response (e.g., chance correct response rates of 50% vs. 25% or 20%). This finding suggests a limited

depth of understanding for many of the issues and that agree-disagree items may overestimate lay knowledge. Further, many researchers relied on items drafted for expert, not lay, respondents. Do lay respondents comprehend the items? As Read and **Desmarais** (2009a) demonstrated, alterations to the wording of items intended to improve comprehensi-bility can improve lay performance. As Hope et al. (2009) suggest, provision of a limited set of response options contribute to an overestimation of knowledge. Finally, we compared lay responses to both question formats with expert responses to agree-disagree items. How would the experts respond to multiple-choice items? How would this change our understanding of expert and lay beliefs regarding eyewitness phenomena? These are empirical questions that remain to be answered.

In contrast with Benton et al.'s (2006) recent assertion regarding the stability of lay knowledge of eyewitness factors, the present data clearly demonstrate improvements in performance over time. This finding has several important implications. First, any evaluation of lay knowledge has a limited shelf life. Citation of research studies from 10, 20, or 30 years ago can provide background to current investigations and their results, but that is all. Introducing the results of older surveys in court as evidence in support of expert eyewitness testimony would provide an inaccurate picture of public beliefs and likely underestimate juror knowledge. Psychological research and our understanding of eyewitness issues have changed dramatically in the last 30 years. As a result, changes in lay knowledge over time should not be unexpected and may reflect successful dissemination of our improved scientific understanding of these phenomena. Changes in the accuracy of lay opinion also correspond with media's increased fascination with criminal investigation procedures and other eyewitness issues over the last 30 years (**Desmarais**, Price, & Read, 2008). Overall, the improvements seen on the eyewitness variables included in this research should provide hope for the future: Our combined efforts in educating the public regarding the pitfalls of eyewitness testimony may finally be having an effect.

There has been considerable discussion in the field around the validity of conducting jury research using undergraduate samples. Bornstein's (1999) informal review of jury decision-making studies revealed few differences between samples of students, community respondents, and actual jurors. Our data, like those of Kassin and Bamdollar (1992), suggest that fears regarding comparability of student and community data may be unwarranted. Specifically, we failed to identify any significant effects of sample type on performance for the 16 items included in this review. In contrast with concerns that student data overestimates lay knowledge of eyewitness issues, when we did see (non-significant) differences between sample types, they often were in the direction of community respondents outperforming students. In other words, our community respondents, who are arguably more likely to serve on actual juries, appear to know more about eyewitness issues than do university students. Those studies that sampled community respondents may still suffer from concerns regarding ecological validity because there may be important demographic and attitudinal differences between community members who do or do not participate in research. Seltzer, Lopes, and Venuti's (1990) method of surveying community members who have participated in trials or Benton et al.'s (2006) sampling of community members who appear for jury duty but are not yet assigned to a case may provide more ecologically valid data. However, legislation in some jurisdictions, including Canada, prohibits jurors from participating in such research. This raises further considerations regarding gen-eralizability because our analyses, although limited in power, identified some significant differences in lay knowledge between jurisdictions.

The ultimate reason for assessing lay knowledge of eyewitness factors is to speak to the common sense criterion of admissibility. Based on this review, what topics are "beyond the ken" of (potential) jurors? The answer depends on our interpretation of the term "common sense." If we define a common sense understanding to require agreement at rates of 80% or greater--the same rate used to establish general agreement amongst experts by Kassin et al. in their 2001 survey--then expert testimony on 12 of the 16 eyewitness factors should not be ruled inadmissible as a matter of course: lineup instructions, mugshot bias, presentation format, accuracy-confidence, child suggestibility, cross-race bias, exposure time, forgetting curve, hypnotic suggestibility, postevent information, unconscious transference, and weapon focus (see weighted means in Table 1). In other words, lay respondents reached consensus only for alcohol intoxication, attitudes & expectations, confidence malleability, and question wording. However, if we define common sense to require agreement at rates of 70% or greater--the same rate used to establish general agreement by Kassin et al. in their 1989 survey--a different picture emerges. Mean correct ***209** responding rates exceeded 70% for more than half of the items: alcohol intoxication, attitudes & expectations, child suggestibility, confidence malleability, lineup instructions, postevent information, and question wording. Finally, if we consider common sense to reflect majority agreement and examine correct responding at rates of 60% or greater, there are only five topics that do not meet this criterion: accuracy-confidence, cross-race bias, hypnotic suggestibility, presentation format, and

weapon focus.

Our attempt to answer this key question regarding lay knowledge leads us to another important one: What is the appropriate standard for determining whether an eyewitness issue is within jurors' common sense understanding? As discussed elsewhere, an empirical examination of judicial beliefs regarding the level of lay knowledge that should be deemed insufficient is likely the only way we will arrive at a resolution regarding the common sense criterion (*cf.* Lane, Groft, & Alonzo, 2008). In cases where jurors must decide certainty of guilt based largely or entirely on eyewitness testimony, it is nonetheless reasonable to suggest that a court be at least 80% certain that jurors accurately comprehend the operation of variables that could affect eyewitness memory accuracy in the case at hand. As reviewed above, use of the 80% criterion would imply that 75% of the eyewitness factors reviewed herein are "beyond the ken" of potential jurors and that, as a result, their decision making may benefit from expert testimony. Ultimately, however, the final decision regarding admissibility will depend on the individual court's interpretation of necessity.

Limitations of the Present Research

A meta-analytic review has inherent limitations. The results are only as good as the data upon which the review is based. Although we used a comprehensive approach to identify both published and unpublished surveys of lay knowledge, it is possible that not all relevant studies were included in the present analyses. By pooling data from different sources, we also unavoidably introduce biases which may limit interpretation of the present findings, such as differences in chance performance rates associated with the various response formats. Further, the data collected and information presented in the primary sources limit our analyses. For instance, researchers differed in the number of items surveyed and the level of detail provided regarding their respondents and results. Further information regarding respondent age, gender, education, and employment, for example, could elucidate inconsistencies across surveys. Finally, we based our analyses on the responses deemed "correct" by the authors of the primary source, but the means through which researchers identified "correct" varied somewhat: some reviewed the empirical literature whereas others relied on expert responding. Although we are confident of significant consistency with regard to the correct response, some differences probably exist across surveys as research findings and expert opinion changed over time.

Conclusions and Future Research

In summary, considering those factors upon which experts themselves have reached consensus, results of this meta-analytic review reveal that there are several factors for which lay knowledge is inadequate. However, there also are many factors upon which lay respondents demonstrate reasonable agreement, across samples, formats, and jurisdictions. Importantly, even if the majority of jurors hold the deemed correct opinion, such opinion may or may not be an adequate safeguard against overbelief of eyewitness evidence. When public beliefs approximate experts' beliefs, lay persons still may not demonstrate the requisite depth of knowledge or sensitivity to the issues in real court cases; that is, whether knowledge regarding eyewitness issues would be applied to the case at hand. In fact, the few studies examining jurors' use of knowledge suggests the answer is "no." For example, Cutler et al. (1989) found that even where jurors had knowledge of the limitations of eyewitness identification, the information was not well-integrated into their decision making. More recently, Martire and Kemp (2009) found similar results. Their research additionally showed that expert eyewitness testimony may not improve juror sensitivity to eyewitness issues.

Our review focused on those eyewitness topics for which the vast majority of researchers believe they know the correct answer. There are many other factors that have been surveyed but were not included in the present review, as well as others to be examined in the future. As a result, we recommend regular re-analysis of the body of work describing lay knowledge of eyewitness issues, as public beliefs, scientific understanding, and expert opinion will change.

<div align="center">

References

References marked with an asterisk indicate studies included in the meta-analysis.

</div>

Benton, T. P., McDonnell, S., Ross, D. F., Thomas, W. N., & Bradshaw, E. (2007). Has eyewitness testimony research penetrated the American legal system? A synthesis of case history, juror knowledge, and expert testimony. In D. F. Ross, R. C. L. Lindsay, J. D. Read, & M. P. Toglia (Eds.), *Handbook of eyewitness psychology: Volume 2: Memory for people* (pp. 453-500). New Jersey: Erlbaum.

*Benton, T. R., Ross, D. F., Bradshaw, E., Thomas, W. N., & Bradshaw, G. S. (2006). Eyewitness memory is still not common *210 sense: Comparing jurors, judges, and law enforcement to eyewitness experts. *Applied Cognitive Psychology, 20,* 115-130. doi:10.1002/acp.1171.

Bornstein, B. H. (1999). The ecological validity of jury simulations: Is the jury still out? *Law and Human Behavior, 23,* 75-92. doi:10.1023/A:1022326807441.

Boyce, M., Beaudry, J., & Lindsay, R. C. L. (2007). Belief of eyewitness identification evidence. In M. P. Toglia, D. F. Ross, J. D. Read, & R. C. L. Lindsay (Eds.), *Handbook of eyewitness psychology: Volume 2: Memory for people* (pp. 501-525). New Jersey: Erlbaum.

Cutler, B. L., & Penrod, S. D. (1995). *Mistaken identification: The eyewitness, psychology, and the law*. New York: Cambridge University Press.

Cutler, B. L., Penrod, S. D., & Dexter, H. R. (1989). The eyewitness, the expert psychologist, and the jury. *Law and Human Behavior, 13,* 311-332. doi:10.1007/BF01067032.

*Deffenbacher, K. A., & Loftus, E. F. (1982). Do jurors share a common understanding concerning eyewitness behavior? *Law and Human Behavior, 6,* 15-30. doi:10.1007/BF01049310.

Desmarais, S. L., Price, H. L., & Read, J. D. (2008). Objection, your honor! Television is not the relevant authority, crime drama portrayals of eyewitness issues. *Psychology, Crime & Law, 14,* 225-243. doi:10.1080/10683160701652583.

* **Desmarais**, S. L., & Read, J. D. (2008). *Re-examining juror knowledge of eyewitness issues*. Unpublished manuscript.

*Devenport, J., & Cutler, B. L. (2009). *Juror knowledge and sensitivity to eyewitness identification testimony in a simulated trial*. Manuscript in preparation.

*Hope, L., Memon, A., Houston, K. A., & Read, J. D. (2009). *Do we need eyewitness expert testimony? A preliminary examination of the methodological considerations in evaluating juror 'commonsense'*. Manuscript in preparation.

*Kassin, S. M. & Barndollar, K. A. (1992). The psychology of eyewitness testimony: A comparison of experts and prospective jurors. *Journal of Applied Social Psychology, 22,* 1241-1249. doi:10.1111/j.l559-1816.1992.tb00948.x.

Kassin, S. M., Ellsworth, P. C, & Smith, V. L. (1989). The "general acceptance" of psychological research on eyewitness testimony: A survey of the experts. *American Psychologist, 44,* 1089-1098. doi:10.1037/0003-066X.44.8.1089.

Kassin, S. M., Tubb, V. A., Hosch, H. M., & Memon, A. (2001). On the "general acceptance" of eyewitness testimony research. *American Psychologist, 56,* 405-116. doi:10.1037//0003-066X.56.5.405.

*Lane, S. M., Graft, S., & Alonzo, J. (2008, March). *What judges know about what jurors know: Consistency of beliefs about eyewitness memory*. Paper presented at the annual meeting of the American Psychology-Law Society, Jacksonville, FL.

Lindsay, R. C. L. (1994). Expectations of eyewitness performance: Jurors' verdicts do not follow from their beliefs. In D. F. Ross, J. D. Read, & M. P. Toglia (Eds.), *Adult eyewitness testimony: Current trends and developments* (pp. 362-384). New York, NY: Cambridge University Press.

Loftus, E. F. (1979). *Eyewitness Testimony*. Cambridge, MA: Harvard University Press.

Martire, K. A., & Kemp, R. I. (2009). The impact of eyewitness expert evidence and judicial instruction on juror ability to evaluate eyewitness testimony. *Law and Human Behavior, 33,* 225-236. doi:10.1007/sl0979-008-9134-z.

*McConkey, K. M., & Roche, S. M. (1989). Knowledge of eyewitness memory. *Australian Psychologist, 24,* 377-384. doi:10.1080/00050068908259576.

*Noon, E., & Hollin, C. R. (1987). Lay knowledge of eyewitness behavior: A British survey. *Applied Cognitive Psychology, 1,* 143-153. doi:10.1002/acp.2350010207.

*Rahaim, G. L., & Brodsky, S. L. (1982). Empirical evidence versus common sense: Juror and lawyer knowledge of eyewitness accuracy. *Law & Psychology Review, 7,* 1-15.

*Read, J. D., & **Desmarais**, S. L. (2009a). Lay knowledge of eyewitness issues: A Canadian evaluation. *Applied*

35 LHUMB 200
35 Law & Hum. Behav. 200

Cognitive Psychology, 32, 301-326. doi:10.1002/acp.1459.

Read, J. D., & **Desmarais**, S. L. (2009b). Expert psychology testimony on eyewitness identification: A matter of common sense? In B. L. Cutler (Ed.), *Expert testimony on the psychology of eyewitness identification* (pp. 115-141). New York: Oxford University Press. doi:10.1093/acprof:oso/9780195331974.003.006.

*Schmechel, R. S., O'Toole, T. P., Easterly, C, & Loftus, E. F. (2006). Beyond the ken? Testing jurors' understanding of eyewitness reliability evidence. *Jurimetrics, 46*, 177-214.

Seltzer, R., Lopes, G. M., & Venuti, M. (1990). Juror ability to recognize the limitations of eyewitness identifications. *Forensic Reports, 3*, 121-137.

Shaw, J. S., Garcia, L. A., & McClure, K. A. (1999). A lay perspective on the accuracy of eyewitness testimony. *Journal of Applied Social Psychology, 29*, 52-71. doi:10.1111/j.1559-1816.1999.tb01374.x.

Wells, G. L. (1978). Applied eyewitness testimony research: System variables versus estimator variables. *Journal of Personality and Social Psychology, 36*, 1546-1557.

Wells, G. L., & Olson, E. A. (2003). Eyewitness testimony. *Annual Review of Psychology, 54*, 277-295. doi:10.1146/annurev.psych.54.101601.145028.

*Yarmey, A. D., & Jones, H. P. T. (1983). Is the psychology of eyewitness identification a matter of common sense? In S. M. A. Lloyd-Bostock & B. R. Clifford (Eds.), *Evaluating witness evidence* (pp. 13-40). Toronto: Wiley.

[FNa1]. S. L. **Desmarais** University of South Florida, 13301 Bruce B. Downs Blvd, MHC 2735, Tampa, FL 33612, USA e-mail: sdesmarais@bcs.usf.edu

[FNa2]. J. Don Read Simon Fraser University, Burnaby, BC, Canada

[FN1]. Although there was a significant statistical association between jurisdiction and year of administration (^{2}[6, N = 22] = 12.84, p < .05, = .75), six cells (75.0%) had an expected count less than five, limiting the stability and practical significance of this finding.

[FN2]. The small number of surveys conducted in the United Kingdom (k = 3) and Australia (k = 2) limited the power of these analyses.

35 Law & Hum. Behav. 200

END OF DOCUMENT

Memory Distortion in Eyewitnesses: A Meta-Analysis of the Post-identification Feedback Effect

AMY BRADFIELD DOUGLASS[1]* and NANCY STEBLAY[2]

[1]*Bates College, USA*
[2]*Augsburg College, USA*

SUMMARY

Feedback administered to eyewitnesses after they make a line-up identification dramatically distorts a wide range of retrospective judgements (e.g. G. L. Wells & A. L. Bradfield, 1998 *Journal of Applied Psychology*, *83*(3), 360–376.). This paper presents a meta-analysis of extant research on post-identification feedback, including 20 experimental tests with over 2400 participant-witnesses. The effect of confirming feedback (i.e. 'Good, you identified the suspect') was robust. Large effect sizes were obtained for most dependent measures, including the key measures of retrospective certainty, view and attention. Smaller effect sizes were obtained for so-called objective measures (e.g. length of time the culprit was in view) and comparisons between disconfirming feedback and control conditions. This meta-analysis demonstrates the reliability and robustness of the post-identification feedback effect. It reinforces recommendations for double-blind testing, recording of eyewitness reports immediately after an identification is made, and reconsideration by court systems of variables currently recommended for consideration in eyewitness evaluations. Copyright © 2006 John Wiley & Sons, Ltd.

Media coverage of DNA exonerations has highlighted the fact that mistaken eyewitness identifications can result in wrongful convictions of innocent suspects (e.g. Doyle, 2005, www.innocenceproject.org). Long before the problem of eyewitness misidentification reached public consciousness, however, psychological researchers explored the memory and social influence processes underlying identification errors. Recently, this research has generated procedures designed to minimize the likelihood of a false identification (Davies & Valentine, 1999; Technical Working Group for Eyewitness Evidence, 1999; Wells, Small, Penrod, Malpass, Fulero, & Brimacombe, 1998). Current recommendations for police lineups include five core components: effective use of fillers (e.g. Wells, Rydell, & Seelau, 1993); blind administration of the line-up (e.g. Douglass, Smith, & Fraser-Thill, 2005; Phillips, McAuliff, Cutler, & Kovera, 1999); a cautionary instruction to the witness that the culprit may or may not be present in the set of photos (Malpass & Devine, 1981; Steblay, 1997), sequential rather than simultaneous presentation of photos (e.g. Lindsay & Wells, 1985; Steblay, Dysart, Fulero, & Lindsay, 2001), and obtaining a statement of certainty from the witness at the time of the identification decision (e.g. Luus & Wells, 1994). As researchers continue to advance knowledge of best line-up practices, a number of

*Correspondence to: A. B. Douglass, Bates College, 4 Andrews Road, Lewiston, ME 04240, USA.
E-mail: adouglas@bates.edu

A. B. Douglass and N. Steblay

jurisdictions in the United States are bringing these science-based recommendations to effective field practice (see Klobuchar, 2005).

Line-up research recently has produced an ancillary line of investigation focusing on the integrity of an eyewitness's recollections *after* the line-up decision is made. This growing body of literature has revealed the astonishing power of a casual comment from a line-up administrator to affect eyewitness memory. The first study to examine this effect (Wells & Bradfield, 1998) demonstrated that confirming post-identification feedback received by the witness immediately after the identification (i.e. 'Good. You identified the actual suspect'.) significantly inflated retrospective confidence reports when compared with a control group told nothing about the accuracy of the identification (participants indicated how certain they were *at the time of their identification*). Perhaps more alarming is that an extensive range of variables was inflated in conjunction with retrospective certainty, including witness reports of the quality of their view of the perpetrator, how much attention was paid, ease of the identification, and basis for the identification. Participants who received confirming feedback were also more willing to testify about their identification and reported a greater ability to remember strangers.

The post-identification feedback effect bears a resemblance to Fischhoff's *hindsight bias* (1977) in which participants given the correct answer to a decision indicated how they would have responded, had they not known the correct answer. Participants' estimates of their own accuracy were routinely higher than the actual accuracy of participants who made the same decision without knowing the correct choice. There are two important differences between Fischhoff's paradigm and the post-identification feedback paradigm. First, in the feedback paradigm participants cannot misremember their prior decision because feedback is administered immediately after the identification is made. Second, participants are asked to recall judgements surrounding a decision, rather than a decision itself (see Bradfield & Wells, 2005, for a fuller discussion of these issues). Therefore, the post-identification feedback effect demonstrates that outcome information can distort memories beyond the boundaries first outlined by Fischhoff.

Subsequent research has replicated the post-identification feedback findings with variations in experimental design designed to explore their theoretical underpinnings. One explanation for the effect hypothesized that participants do not consider their judgements before being queried in the dependent measures questionnaire. At that time, the only way to consider judgements about the witnessed event and the identification procedure is through the lens of the feedback received. One set of experiments explicitly tested this possibility. Wells and Bradfield (1999) manipulated whether participants were instructed to think about testimony-relevant judgements before receiving feedback. Participants who answered questions about their certainty before hearing feedback were inoculated against the effect of feedback on the certainty dependent measure—their judgements did not show the typical post-identification feedback inflation on retrospective certainty. Research in a related paradigm demonstrated a similar ability of prior thought to protect participants against the memory distorting effects of feedback (Bradfield & Wells, 2005).

The post-identification feedback effect is noteworthy for multiple reasons. First, eyewitnesses in the feedback paradigm typically have made identifications from target-absent photospreads—all of their identifications are inaccurate. Consequently, their distorted reports correspond to mistaken identifications of innocent suspects, a forensically relevant scenario of critical importance given the eyewitness errors exposed by DNA exoneration cases (e.g. Davies, 1996; Rattner, 1988). Second, this powerful effect is

produced by a simple and seemingly casual comment from the line-up administrator—a 'system' variable (Wells, 1978) that potentially could be controlled in police practice. Third, the aspects of eyewitness experience distorted by post-identification feedback (e.g. certainty, witness perception of his/her view of the perpetrator, attention given to the witnessed event, ease of identification) are the very attributes that are likely to bolster eyewitness credibility in the eyes of investigators, prosecutors, and juries. Research has established that people who evaluate eyewitness identifications routinely and naturally assume that confidence (certainty) is correlated with accuracy (e.g. Leippe, 1994) and continue to use confidence to assess accuracy even after being told that the two are not reliably linked (Fox & Walters, 1986). Finally, court systems have explicitly recommended using some of the very criteria distorted by post-identification feedback in evaluations of eyewitnesses. The US Supreme Court recommends using certainty, view, and attention reports (e.g. *Neil v. Biggers*, 1972); courts in England and Wales recommend using view and attention (*R v Turnbull*, 1977); the Australian Law Reform Commission (2005) is currently reviewing jury instructions regarding eyewitness identification evidence. Considering the research findings reported above, these recommendations demand scrutiny.

Since the post-identification feedback effect entered the published eyewitness literature in 1998, many researchers have explored this phenomenon. However, in spite of strong academic interest in the topic, there has not been a systematic organization and evaluation of the research. The current research aims to provide such structure using the tool of meta-analysis. Meta-analysis already has been useful in the psycho-legal realm, as it provides objective quantitative indicators of the status of a hypothesized effect, detailed analysis of effect moderators and direction for advances in research design, theory and practice (see recent meta-analyses on topics of line-up instruction, Steblay, 1997; sequential presentation, Steblay, et al., 2001; and showups, Steblay, Dysart, Fulero, & Lindsay, 2003.).

Systematic evaluation of extant research on the post-identification feedback effect will assist future researchers by guiding the selection of variables and experimental paradigms that can target the causes and parameters of the effect. Equally important, this meta-analysis will provide a summary of knowledge for a broader audience that includes line-up administrators and court personnel. Line-up administrators are often interested in learning about strategies for obtaining eyewitness evidence that are immune to challenges from the defence. Similarly, court personnel (including defence attorneys) are interested in hearing from experts about procedures that might have compromised the integrity of the eyewitness's memory. With a meta-analysis as a foundation for their recommendations, experts involved in conversations with these constituencies will be able to inform both more comprehensively.

METHOD

Sample

The sample included 20 experimental tests from 14 studies. Studies were obtained from a search of PsycInfo and additional conversation with researchers within this area of expertise. The final sample included 10 published and 4 unpublished studies, representing 2477 participant-witnesses. The majority of the studies were conducted in the United

A. B. Douglass and N. Steblay

States (n = 11) with others conducted in the United Kingdom (n = 1), and Australia (n = 2). In order to be included in the meta-analysis, the study must have included a laboratory test of the confirming feedback effect, retrospective certainty as a dependent variable, and data that could provide calculation of an effect size for the comparison of a group that received confirming feedback to a control group.

The studies included in this analysis were conducted between 1998 and 2005 using participants who ranged in age from 11 to 97; most were college students. Participants of both genders were included in 100% of the studies. All studies used videotaped stimuli as the witnessed event with a range of length from 60 seconds to 180 seconds and required participants to make an identification from a photospread containing colour photographs. One exception was Bradfield, Wells, & Olson, 2002 in which participants made an identification from a videotaped lineup. Sample sizes ranged from 62 to 320 ($M = 176.93$).

Dependent measures

A total of 13 dependent measures were recorded from the 14 studies analysed, not all of which were included in each study analysed. The measures fell into three broad categories. First were retrospective judgements regarding the witnessed event. Measures in this category include: *view, attention paid, ability to make outfacial features, basis for an identification, quality of the culprit's image in memory, distance of the camera from the perpetrator*, and *length of time the perpetrator wasin view*. A second set of measures concerned aspects of participants' identification experience: *retrospective certainty, easeof identification* and *timeneeded to make the identification*. Finally, measures concerning summative judgements were analysed: *general ability to remember strangers, reports of trust in an eyewitness who had similar viewing conditions*, and *willingness to testify*.

Statistics

Cohen's d, the standardized mean difference between two groups, was calculated as the effect size indicator for each comparison (Cohen, 1988). In the following results, d is used to indicate a mean effect size across tests. A meta-analytic Z (Z_{ma}) was calculated using Rosenthal's (1991) method of combining t-values. This Z_{ma} provides an overall probability level associated with the observed pattern of results. A fail-safe N (N_{fs}) was calculated as a means to determine the number of unretrieved studies averaging null results necessary to bring the overall *p*-value to a specific level of significance (in this case, p = 0.05). This number of studies, or tolerance for future null results, allows us to evaluate the resistance of the review conclusion to a 'file drawer threat' (Rosenthal, 1991).

Comparisons

Eleven of the tests compared a confirming feedback (CF) condition to a no feedback (NF) control group. Six compared CF to disconfirming feedback (DF) condition. Three compared DF and NF groups. The focus of our study is the first comparison (CF vs. NF) as that is the forensically relevant contrast because of the inflationary power of confirming feedback for a witness who has identified a suspect.

RESULTS

Primary analysis: Comparison between confirming feedback (CF) and no feedback (control) groups, on each dependent measure

Certainty

Certainty is arguably the most important dependent measure in the post-identification feedback paradigm. Participants receiving confirming feedback expressed significantly more retrospective confidence in their decision compared with participants who received no feedback ($d = 0.79$, $Z_{ma} = 13.42$, $p < 0.0001$). An effect size of 0.79 is considered large, based on Cohen's rule of thumb (Cohen, 1988) (see Table 1).

Biggers criteria

Eyewitness certainty (noted above), opportunity to view the perpetrator, and attention paid to the event are qualities of the eyewitness viewing experience that, according to the US Supreme Court (*Neil v. Biggers*, 1972), are criteria relevant to juror decision-making (recommendations in England and Wales focus on view and attention). Participants' retrospective reports of view and attention ($ds = 0.50$ and 0.46, respectively) were significantly affected by confirming feedback, producing medium effect sizes.

Related subjective measures

CF participants demonstrated consistently inflated perceptions on subjective measures related to their line-up performance compared to the control group. Participants who received confirming feedback reported that they possessed a significantly better basis for making the identification ($d = 0.77$), greater clarity of the perpetrator's image in mind ($d = 0.68$), greater ease of identification ($d = 0.80$), and needing less time to make their ID ($d = 0.45$). They also reported a better memory for strangers' faces ($d = 0.45$) and greater trust in the memory of another witness with a similar experience ($d = 0.52$). Not surprisingly, then, they are also more willing to testify about their identification decision ($d = 0.82$).

Table 1. Confirming feedback vs. No feedback comparison: Retrospective reports

Dependent measure	Tests	d	Range	(min, max)	Nfs
Certainty at time of ID	11	0.79	0.20	1.27	590
How good a view?	9	0.50	−0.02	0.90	132
Opportunity to view face	9	0.55	−0.04	1.04	165
Attention paid	9	0.46	0.27	0.67	145
Good basis to make an ID	9	0.77	0.56	1.10	386
Ease of making an ID	9	0.80	0.35	1.02	587
Speed of ID	9	0.45	0.12	0.67	104
Willing to testify	9	0.82	0.43	1.13	437
My memory for strangers	8	0.45	0.19	0.84	75
Clarity of image in my mind	7	0.68	0.30	1.17	150
Trust in eyewitness with similar experience	3	0.52	0.41	0.71	<1
How far away?	2	0.12	0.10	0.13	<1
How long in view?	4	0.29	0.09	0.69	<1
Confidence 'right now'	2	0.53	0.31	0.75	19

Comparisons for all measures produced statistically significant Z_{ma} values ($p < 0.05$), *except* for 'how far away' and 'how long in view'.

A. B. Douglass and N. Steblay

'Objective' measures
Smaller effect sizes and no statistically significant differences were found for attributes of
the participants' experience that are ostensibly objective: time that the perpetrator was in
view and distance from the camera to the perpetrator (ds = 0.29, and 0.12, respectively).
These smaller effect sizes are noteworthy for at least two reasons: They suggest some
limits to the influences of confirming feedback; they also indicate discernment and
seriousness on the part of the research participants (i.e., participants were not simply
employing a thoughtless response set across measures). However, they also tap information
to which a subject could surmise the experimenter has access—knowable facts—unlike an
investigator in a real crime situation.

Moderators
The small number of studies available did not allow for extensive moderator analysis.
However, it may be noted that post-identification feedback effects are quite robust. Overall,
the studies involved a reasonably diverse sample of participants (undergraduates, children,
adults) and stimulus materials. The effects were achieved for witnesses who made accurate
identifications in target-present lineups as well as false IDs in target absent arrays, although
the effect is stronger for inaccurate witnesses (Bradfield et al., 2002). Semmler, Brewer, &
Wells (2004) found post-identification effects for witnesses who rejected the lineup also
('He's not there'). Semmler et al., also found the effects when a cautionary instruction
('may or may not be in the lineup') was provided; Douglass and McQuiston-Surrett
(in press) found the effects with both sequential and simultaneous lineups.

**Secondary analysis: Comparison between disconfirming feedback (DF)
and no feedback (Control) groups, on each dependent measure.**

Only three tests (in three separate studies) explored the impact of disconfirming feedback
on participants' retrospective reports. These reports produce small average effect sizes and
some inconsistencies. For dependent measures of *view* and *ease of ID*, participants
receiving disconfirming feedback indicate less positive retrospective reports in all three
studies, (d = −0.14 and −0.31, respectively). On measures of retrospective confidence
(d = −0.21), ability to make out details of the face (d = −0.04), attention (d = −0.08),
basis for ID (d = −0.10), time to make an ID (d = 0.01), and willingness to testify
(d = −0.10), the three tests show mixed results—two tests with less positive reports from
the DF condition, one test with more positive reports (negative effect sizes indicate higher
scores from the NF control condition).[1]

DISCUSSION

Through this review, the reliability and robustness of the post-identification feedback effect
are well documented. Over 2400 participant-witnesses have been tested, with remarkably
consistent outcomes. Compared to control participants, those who receive a simple post-
identification confirmation regarding the accuracy of their identification significantly

---

[1]Not surprisingly, a comparison of CF and DF conditions indicates substantial differences between the groups in
retrospective certainty (d = 1.07), view (0.69), memory for the face (0.77), attention (0.67), basis for judgment
(0.79), ease of ID (1.01), time to make an ID (0.60), willingness to testify (1.01), memory for strangers (0.58),
clarity of image in memory (0.99), and trust in an eyewitness with similar experience (0.56).

inflate their reports to suggest better witnessing conditions at the time of the crime, stronger memory at the time of the lineup, and sharper memory abilities in general. Participants apparently make what would otherwise seem to be reasonable *post hoc* inferences about their witnessing experience and behaviour during the identification. However, these inferences are based on erroneous information external to their actual memory of these events. This 'creeping determinism' (Fischhoff, 1975) produces a memory distortion that is by no means benign.

The implications of these results are quite profound. Both memory for a crime and confidence in one's memory are fragile and potentially slippery evidentiary elements. Indeed, one of the startling lessons of line-up research is just how powerful seemingly subtle aspects of line-up construction and investigator behaviour can be. The simple addition of a cautionary instruction (the perpetrator 'may or may not be in the lineup') produces a significant (25%) drop in false identifications (Steblay, 1997); and use of a sequential lineup cuts the false identification rate almost in half (by 23%) in target absent lineups (Steblay et al., 2001). Similarly, subtle changes in investigator behaviour derived from the knowledge an investigator has about the identity of the suspect can influence witnesses' identification decisions (e.g. Douglass et al., 2005; Phillips et al., 1999) and confidence (Garrioch & Brimacombe, 2001).

Although the present data do not allow for a precise calculation of the number of errors that could be avoided with a change in practice, they do provide dramatic evidence that post-identification feedback can compromise the integrity of a witness's memory. Wells and Bradfield (1998) made this point clearly when reporting participant-witnesses' disproportionate use of the extreme end of the 'certainty' scale: 50% of CF participants in that study indicated certainty of six or seven on a 7-point scale (compared with 15% using the extreme end of the scale in the DF condition). The relevance for real cases is clear. First, as noted by Wells and Bradfield (1998), it is reasonable to assume that an eyewitness must exceed some threshold of credibility in order for investigators and prosecutors to move ahead in their case against a suspect. Witnesses who reconstruct and enhance their report of both witnessing and identification procedures may well increase the likelihood that a case against that suspect will be pursued. Frighteningly, this enhancement is not due to increased accuracy, but to extra-memory factors. Second, witnesses with feedback-enhanced memories will likely be more compelling witnesses at trial, increasing the chances of a conviction—an unwelcome outcome if an innocent suspect was identified.

Clear understanding of the impact of post-identification confirmation can facilitate the goal of many eyewitness researchers—prevent mistaken identifications from resulting in wrongful convictions. This meta-analysis can help accomplish this goal in several ways. First, this research should provide police with a strong rationale as to why it is critically important to administer double-blind photospreads and to immediately record eyewitness confidence. These procedures could decrease the likelihood that juries will be erroneously impressed by a falsely confident eyewitness. This is especially critical because at least one study demonstrates that participant-jurors are not sensitive to eyewitnesses who display confidence that has inflated over time (Bradfield & McQuiston, 2004). Additionally, this meta-analysis should influence the treatment of information regarding post-identification feedback effects in court by providing attorneys and experts with a stronger foundation from which to argue that a witness's memory could be distorted if double-blind procedures were not followed and immediate confidence reports not recorded. For experts who testify in court, this meta-analysis will facilitate admittance of testimony on this topic. Most American courts now use the Daubert standard for admitting expert testimony, one element

A. B. Douglass and N. Steblay

of which is that the information presented by an expert must have achieved 'general acceptance in the relevant scientific community' (*Daubert v. Merrell Dow Pharmaceuticals*, 1993). The consistency in outcomes demonstrated in this review lends credence to the argument that post-identification feedback effects should be 'generally accepted''.

Directions for future research

The effects reported here are remarkably consistent, suggesting that future research will target explanations for the post-identification feedback effect rather than resolution of inconsistencies. One direction for future research is to identify ways in which to anchor the witness's memory in the witnessing experience itself rather than in post-event information. Wells and Bradfield (1999) have found some success in moderating the post-identification effect with instructions to the witness to privately think about his or her confidence and attributes of the witnessing experience prior to receiving feedback. The videotaping of line-up procedures may also provide the means to later remind a witness (as well as a jury) of his or her confidence and perceptions at the time of the lineup (e.g. Kassin, 1998; Sporer, 1993). However, because even a 48-hour delay did not diminish distorted retrospective reports (Wells, Olson, & Charman, 2003), finding other ways to anchor witnesses' memory is critical.

Although the small number of studies did not allow for comprehensive analyses of moderator variables, there was enough evidence to suggest that 'objective' measures are less susceptible to memory distortion than are 'subjective' measures. Therefore, it might be worthwhile to examine this variable more systematically. Perhaps the difference is due to the fact that participants realize those questions can be evaluated for accuracy by the experimenters. Would the same difference appear in a paradigm where participants knew that experimenters did not have access to accurate answers (i.e. in a more ecologically valid paradigm)? Other directions for future research include pursuing explanations for conditions under which the feedback effect is diminished such as when disconfirming feedback is administered. Although witnesses who receive disconfirming feedback probably have minimal impact on the criminal justice system (i.e. because they do not testify in court), an explanation for the smaller effects of disconfirming feedback could provide information about the nature of eyewitness memory and how it interacts with social influence cues. Finally, researchers might pursue feedback analogues. Would learning about accuracy from sources outside the immediate identification experience—e.g. a news report, a prosecutor, another witness (cf. Luus & Wells, 1994)—have the same distorting effects on retrospective confidence and perceptions of the witnessing conditions? The current research reveals an increased willingness of witnesses to testify in court. Does this eagerness translate to differences in subsequent interview and/or courtroom behaviour?

Recommendations

The primary recommendation to be made from this meta-analysis is straightforward—feedback to the witness should not be part of the identification procedure. There is also a straightforward strategic solution: use a blind line-up administrator, thoroughly record the line-up process, and obtain eyewitness reports (particularly confidence) immediately after the identification. Currently, blind administrators are recommended in order to guard against memory errors during the line-up decision (e.g. Wells et al., 1998). In England and Wales, the Police and Criminal Evidence Act (PACE) dictates that an officer who is not

involved in the investigation conduct the line-up procedure (although he or she does know who the suspect is, Davies & Valentine, 1999). Collateral benefits of a blind administrator are afforded in that no feedback could be provided to the witness until after completion of the line-up procedure and documentation of testimony-relevant judgements (Technical Working Group for Eyewitness Evidence, 1999).

A final recommendation is directed at courts considering recommendations for evaluations of eyewitnesses. In the United States, the Supreme Court should reconsider its current recommendation for evaluations of eyewitness testimony. Of the five criteria outlined by the US Supreme Court in *Neil v. Biggers* (1972), three are dramatically distorted by post-identification feedback: confidence, attention and view. Similarly, the Turnbull Rules in England and Wales also include two variables distorted by post-identification feedback: attention and view (*R v Turnbull*, 1977). Rulings from courts suggesting that these variables only be taken into account if post-identification feedback has not been administered would likely do much to decrease the incidence of feedback in real world cases. Barring reconsideration of these criteria in court recommendations, researchers should continue to press for immediate witness reports and blind line-up administration. These practices are best suited to prevent the memory distorting effects of post-identification feedback.

ACKNOWLEDGEMENTS

REFERENCES

*Articles included in the meta-analysis

Australian Law Reform Commission (2005). Discussion Paper 69, Review of the Uniform Evidence Acts. Retrieved January 4, 2006 from http://www.austlii.edu.au/au/other /alrc/publications/dp/69/
Bradfield, A., & McQuiston, D. E. (2004). When does evidence of eyewitness confidence inflation affect judgments in a criminal trial? *Law and Human Behavior, 28*(4), 369–387.
Bradfield, A., & Wells, G. L. (2005). Not the same old hindsight bias: Outcome information distorts a broad range of recollections. *Memory and Cognition, 33*(1), 120–130.
*Bradfield, A. L., Wells, G. L., & Olson, E. A. (2002). The damaging effect of confirming feedback on the relation between eyewitness certainty and identification accuracy. *Journal of Applied Psychology, 87*, 112–120.
*Coddington, K. A., & Brigham, J. C. (2000, March). The malleability of eyewitness metamemory judgments: The effect of question difficulty. Poster presented at the biennial meeting of the American Psychology-Law Society meeting. New Orleans, LA.
Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
Daubert v. Merrell Dow Pharmaceuticals, Inc.,113 S.Ct. 2786 (1993).
Davies, G. M. (1996). Mistaken identifications: Where law meets psychology head on. *The Howard Journal, 35*, 232–241.
Davies, G. M., & Valentine, T. (1999). Codes of practice for identification. *Expert Evidence: International Journal of Behavioural Sciences in Legal Context, 7*(1), 59–65.
*Douglass, A. B. (see Bradfield), & McQuiston-Surrett, D. (in press). Post-identification feedback: Exploring the effects of sequential photospreads and eyewitnesses' awareness of the identification task. *Applied Cognitive Psychology*. DOI: 10.1002/acp1253

Douglass, A. B. (see Bradfield), Smith, C., & Fraser-Thill, R. (2005). A problem with double-blind photospread procedures: Photospread administrators use one eyewitness's confidence to influence the identification of another eyewitness. *Law and Human Behavior, 29*(5), 543–562.

Doyle, J. M. (2005). True witness: Cops, courts, science, and the battle against misidentification. Palgrave Macmillin: New York.

Fischhoff, B. (1975). Hindsight ≠ foresight. *Journal of Experimental Psychology: Human Perception & Performance, 1*, 288–299.

Fischhoff, B. (1977). Perceived informativeness of facts. *Journal of Experimental Psychology: Human Perception & Performance, 3*, 349–358.

Fox, S. G., & Walters, H. A. (1986). The impact of general versus specific expert testimony and eyewitness confidence upon mock juror judgment. *Law and Human Behavior, 10*, 215–228.

Garrioch, L., & Brimacombe, C. A. E. (2001). Lineup administrators' expectations: Their impact on eyewitness confidence. *Law and Human Behavior, 25*(3), 299–315.

*Hafstad, G. S., Memon, A., & Logie, R. (2004). Post-identification feedback, confidence and recollections of witnessing conditions in child witnesses. *Applied Cognitive Psychology, 18*, 901–912.

Kassin, S. (1998). Eyewitness identification procedures: The fifth rule. *Law and Human Behavior, 22*, 649–654.

Klobuchar, A., Steblay, N., & Caligiuri, H. (in press). Improving eyewitness identifications: Hennepin County's blind sequential lineup pilot project. *Cardozo Public Law, Policy & Ethics Journal.*

Leippe, M. R. (1994). The appraisal of eyewitness testimony. In D. F. Ross, J. D. Read, & H. P. Toglia (Eds.), *Adult eyewitness testimony: Current trends and developments* (pp. 385–418). New York: Cambridge University Press.

Lindsay, R. C. L., & Wells, G. L. (1985). Improving eyewitness identifications from lineups: Simultaneous versus sequential lineup presentation. *Journal of Applied Psychology, 66*, 79–89.

Luus, C. A. E., & Wells, G. L. (1994). The malleability of eyewitness confidence: Co-witness and perseverance effects. *Journal of Applied Psychology, 79*, 714–723.

Malpass, R. S., & Devine, P. G. (1981). Eyewitness identification: Lineup instructions and the absence of the offender. *Journal of Applied Psychology, 79*, 714–734.

Neil v. Biggers. 409 U.S. 188 (1972).

*Neuschatz, J. S., Preston, E. L., Burkett, A. D., Toglia, M. P., Lampinen, J. M., Neuschatz, J. S., Fairless, A. H., Lawson, D. S., Powers, R. A., Goodsell, C. A. (2005). The effects of post-identification feedback and age on retrospective eyewitness memory. *Applied Cognitive Psychology, 19*, 435–453.

Phillips, M. R., McAuliff, B. D., Kovera, M. B., & Cutler, B. L. (1999). Double-blind photoarray administration as a safeguard against investigator bias. *Journal of Applied Psychology, 84*, 940–951.

R v. Turnbull. Queen's Bench 224 (1977).

Rattner, A. (1988). Convicted but innocent: Wrongful conviction and the criminal justice system. *Law and Human Behavior, 12*(3), 283–293.

Rosenthal, R. (1991). *Meta-analytic procedures for social research.* Newbury Park, CA: Sage.

*Semmler, C., Brewer, N., & Wells, G. L. (2004), Effects of postidentification feedback on eyewitness identification and nonidentification confidence. *Journal of Applied Psychology, 89*(2), 334–346.

Sporer, S. L. (1993). Eyewitness identification accuracy, confidence and decision times in simultaneous and sequential lineups. *Journal of Applied Psychology, 78*(1), 22–33.

Steblay, N. (1997). Social influence in eyewitness recall: A meta-analytic review of lineup instruction effects. *Law and Human Behavior, 21*(3), 283–298.

Steblay, N., Dysart, J., Fulero, S., & Lindsay, R. C. L. (2001). Eyewitness accuracy rates in sequential and simultaneous lineup presentations: A meta-analytic comparison. *Law and Human Behavior, 25*(5), 459.

Steblay, N., Dysart, J., Fulero, S., & Lindsay, R. C. L. (2003). Eyewitness accuracy rates in showup and lineup presentations: A meta-analytic comparison. *Law and Human Behavior, 27*(5), 523–540.

Technical Working Group for Eyewitness Evidence. (1999). *Eyewitness evidence: A guide for law enforcement.* Washington, DC: National Institute of Justice; NCJ 178240.

Wells, G. L. (1978). Applied eyewitness-testimony research: System variables and estimator variables. *Journal of Personality & Social Psychology, 36*(12), 1546–1557.

*Wells, G. L., & Bradfield, A. L. (1998). "Good, you identified the suspect": Feedback to eyewitnesses distorts their reports of the witnessing experience. *Journal of Applied Psychology, 83*(3), 360–376.

*Wells, G. L., & Bradfield, A. L. (1999). Distortions in eyewitness' recollections: Can the postidentification-feedback effect be moderated? *Psychological Science, 10*, 138–144.

*Wells, G. L., Olson, E. A., & Charman, S. D. (2003). Distorted retrospective eyewitness reports as functions of feedback and delay. *Journal of Experimental Psychology: Applied, 9*, 42–52.

Wells, G. L., Rydell, S. M., & Seelau, E. P. (1993). The selection of distractors for eyewitness lineups. *Journal of Applied Psychology, 78*(5), 835–844.

Wells, G. L., Small, M., Penrod, S., Malpass, R. S., Fulero, S. M., & Brimacombe, C. A. E. (1998). Eyewitness identification procedures: Recommendations for lineups and photospreads. *Law and Human Behavior, 22*, 603–647.

Show-ups: The Critical Issue of Clothing Bias

JENNIFER E. DYSART[1]*, R. C. L. LINDSAY[2] and PAUL R. DUPUIS[3]

[1]*John Jay College of Criminal Justice, City University of New York, New York, NY, USA*
[2]*Queen's University, Kingston, Ontario, Canada*
[3]*Red Deer College, Red Deer Alberta, Canada*

SUMMARY

A field study (N = 379) investigated the effects of clothing bias on show-up identifications using variations in type of clothing (distinct and common), the similarity of clothing between the event and the identification procedure, target-present and two target-absent show-ups (high similarity and low similarity innocent suspects), and time delay. Results showed a significant clothing bias by clothing type interaction on identification accuracy; however, no overall effects of delay or common clothing on identification accuracy were found. With distinct clothing, significant effects of clothing bias and suspect similarity emerged. Implications for police use of show-ups are discussed. Copyright © 2006 John Wiley & Sons, Ltd.

In 1990, witnesses to a gang shooting in New Jersey described one of the gang members as wearing a Day-Glo reversible jacket (orange and burgundy) with the orange side turned out. Shortly after the shooting, the police apprehended Luis Kevin Rojas wearing an orange and burgundy reversible jacket, with the burgundy side out. After being asked to turn the jacket inside out to expose the orange side, Mr. Rojas was identified by seven eyewitnesses as being a gang member who had been involved in the shooting (Flynn, 1998). Was Mr. Rojas identified because he was presented to the witness in a show-up rather than a lineup? Would he have been identified from the show-up even if he had not been forced to wear his coat orange side out? In other words, was this a mistaken identification resulting from accumulated biases? A brief review of the show-up literature will demonstrate that show-ups are a risky method of eyewitness identification. Further, a review of the (very limited) literature suggests that clothing bias is a serious concern. We will argue that the show-up is an identification procedure prone to clothing bias and then explore the extent and some limiting conditions, of the clothing bias effect on show-ups in a field study.

The show-up involves exposing a witness to one person, the suspect, either live or in a photograph. A potential problem with this procedure is that with only one person shown to the eyewitness, the identity of the police suspect is obvious. In fact, when the identity of a suspect is obvious in a lineup, the lineup is considered unfair (e.g. Lindsay & Wells, 1980; Phillips, McAuliff, Kovera, & Cutler, 1999). Thus, the inherent suggestiveness of the show-up may lead a witness to identify an innocent suspect as the perpetrator. Despite the

suggestiveness of the show-up, past research on show-up identifications has reported a reduced tendency for witnesses to choose anyone from a show-up, including the perpetrator (Gonzalez, Ellsworth, & Pembroke, 1993; Wagenaar & Veefkind, 1992). In addition, a recent meta-analysis of show-up studies found that show-ups and simultaneous lineups result in virtually identical rates of correct identification (Steblay, Dysart, Fulero, & Lindsay, 2003). The fact that guilty suspects are no more likely to be selected from show-ups than lineups is puzzling. The complete absence of foils (or distractors) in the show-up would seem to make the task easier for the witness and thus, one would expect the show-up to result in higher correct identification rates of the perpetrator. One possible explanation for this finding is that witnesses respond to the perceived pressure to choose with psychological reactance (Brehm & Brehm, 1981) and employ a more stringent decision criterion. Regardless of the reason, show-ups appear to provide neither an advantage nor disadvantage over lineups in terms of correct identifications.

Historically, the concerns expressed about show-ups have centred on the potential for mistaken identification. Some studies support biasing effects of show-ups (e.g. Dysart & Lindsay, in press; Yarmey, Yarmey, & Yarmey, 1996) but others report false identification rates from show-ups that are comparable to those from simultaneous lineups (Lindsay, Pozzulo, Craig, Lee, & Corber, 1997). Finally, some studies have found that show-ups yield more accurate identification decisions overall than lineups (Dekle, Beal, Elliott, & Huneycutt, 1996; Yarmey, Yarmey, & Yarmey, 1994). This last finding, however, can be misleading. Show-ups generate higher rates of correct decisions because they produce higher rates of correct rejection in target-absent conditions as compared to simultaneous lineups. The concern with show-ups, however, is that they do not afford the same protection to innocent suspects, that is provided by using foils (distracters) in lineups. In lineups, incorrect foil selections are known errors and do not result in any legal action against the foil that is selected. With show-ups, all false positive choices are false identifications of innocent suspects. In support of this concern, the Steblay et al. (2003) meta-analysis indicated that false identifications of the suspect were substantially more likely from show-ups than lineups.

Eyewitness research may also be underestimating the rate of false identifications with show-ups in the real world. The reason for this is because lineups usually require that police have a specific suspect in custody and an arrest usually requires some source of evidence leading police to have probable cause to believe the suspect is guilty prior to the identification procedure. Show-ups, however, frequently are based on minimal information, usually that the suspect resembles a description provided by a witness and that the suspect was apprehended near the scene of the crime within a generally brief period following the crime. Given that eyewitness descriptions are frequently vague (Lindsay, Nosworthy, Martin, & Martynuck, 1994), show-ups may commonly contain innocent suspects. Furthermore, police and the courts justify the use of show-ups, at least in part, by arguing that they provide a means of quickly and efficiently exonerating the innocent. If this logic is used, police would be expected to have a more lenient criterion for presenting suspects in show-ups. As a result, the base rate of suspect innocence may be substantially higher for show-ups than lineups. Combined with an equal or higher propensity for false identification, a higher base rate of innocence could lead to a much higher rate of false identification from show-ups than lineups.

Despite the intuitive and empirically demonstrated problems with the show-up procedure described thus far, it is a common method of identification used by police in the United States. Flowe, Ebbesen, Burke, and Chivabunditt (2001) reported that 55% of

identifications conducted in 488 cases between 1991 and 1995 in a large metropolitan area in the western US used show-ups. McQuiston and Malpass (2001) reported that show-ups were used by police in El Paso County, Texas, for 30% of identification attempts. Gonzalez et al. (1993) enlisted the help of a detective in Illinois to record all identifications (line-ups and show-ups) he was involved in for a period of time. Results from this field study indicated that 77% of identifications conducted were show-ups. Because of these sizeable percentages, and the limited number of studies that have investigated factors that may affect the show-up procedure, more studies are needed which demonstrate the potential limitations or advantages of this identification technique.

Some researchers have attributed lineup identification errors to the use of relative judgement strategies (i.e. the witness treating the lineup like a multiple choice test and choosing the lineup member most similar to the memory trace of the criminal) (Lindsay & Bellinger, 1999; Lindsay & Wells, 1985; Wells, 1984; for an alternative view see Clark & Davey, 2005). Lineup biases may function by making the suspect easy to select relative to other lineup members (Lindsay, Wallbridge, & Drennan, 1987; Lindsay & Wells, 1980) or by encouraging witnesses to employ relative judgements (Malpass & Devine, 1981). Sequential lineups have been shown to reduce reliance on relative judgements (Lindsay & Bellinger, 1999) and consistently reduce false positive choices of innocent suspects (Steblay, Dysart, Fulero, & Lindsay, 2001). Lineup members in sequential lineups are presented individually to prevent witnesses from using a comparative strategy. Since, by definition, show-ups involve exposing the witness to a single person, relative judgements are not possible with show-ups. As a result, the show-up procedure may produce fewer false identifications than what would be predicted based on the suggestive nature of the procedure.

Lineup biases have been shown to have less impact with sequential lineups (Lindsay, Lea, & Fulford, 1991; Lindsay, Lea, Nosworthy, Fulford, Hector, LeVan, & Seabrook, 1991). Clothing bias, where only the suspect appears in the identification procedure wearing clothing similar to that described by the eyewitness, has been shown to increase false identifications from lineups (Lindsay & Wells, 1980; Lindsay et al., 1987; Malpass & Devine, 1981), mug shots (Lindsay et al., 1994), and show-ups (Yarmey et al., 1996). However, the impact of clothing bias has not sufficiently been tested in association with show-ups.

Clothing bias is of particular concern with the show-up procedure because suspects presented in show-ups often are apprehended because they were near the scene of the crime and matched the overall description given by the eyewitness. This concern is supported by the fact that approximately 50% of all information reported in eyewitness descriptions is clothing information (Lindsay et al., 1987, 1994). Furthermore, clothing descriptions may be more distinctive than person descriptions. If a witness describes a robber as a clean-shaven white male in his early 20s, the police are unlikely to stop everyone they see on the street who fits that description. However, if the description also includes information that the robber was wearing blue jeans, a white T-shirt, and a baseball cap, police may rely heavily on these clothing cues to obtain a suspect near the scene of the crime who fits the description of the clothing and is not inconsistent with the vague person description. This scenario, which is likely to be common with show-up identifications, leads to potential clothing bias because the suspect is likely to be wearing clothing similar to that remembered by the eyewitness.

To date, one study (Yarmey et al., 1996) has investigated the effects of clothing bias on identification accuracy from show-ups. Yarmey et al. found that when a similar looking

(innocent) suspect was presented in a show-up wearing the same clothing that was worn by the target, false identifications were higher than when the innocent suspects wore different clothing than that worn by the target during the event. Yarmey et al. also presented participants with implausible innocent suspects who did not match the description of the targets in any manner. In this condition, clothing bias had no significant effect on the false identification rate. In addition, Yarmey et al. found that correct identifications of the target were not affected by the clothing manipulation, supporting the notion that clothing bias only serves to decrease identification accuracy (Lindsay et al., 1987).

There are many unanswered questions with respect to clothing bias that need to be addressed, particularly when combined with the show-up procedure. One question relates to the probability that a similar looking innocent person, wearing exactly the same clothing as the perpetrator, will be in the vicinity of the crime immediately after the event. Although this scenario is possible, it seems rather unlikely and, thus, the manipulation used by Yarmey et al. (1996), where the innocent person both resembled the target (her sister) and was presented in the exact clothing worn by the perpetrator, may not accurately reflect most potentially dangerous clothing bias situations in the real-world. Addressing this issue with lineups, Lindsay et al. (1987) investigated the effects of identical, similar, and different clothing on identification accuracy. There was no difference in the impact of identical and similar clothing, both produced the clothing bias effect. Clothing similarity was enough to make the innocent target stand out in the lineup and result in selections of him. The current study investigated the effects of dissimilar, similar, and identical clothing on identification accuracy from show-ups.

A second question that has yet to be addressed in clothing bias investigations is whether the type of clothing worn by the perpetrator has an effect on the magnitude of the biasing effect. For example, if a perpetrator is described as wearing a white T-shirt, is a witness more or less likely to be affected by clothing bias than if the perpetrator was described as wearing a distinctive shirt (e.g. a shirt with a distinctive logo, photograph, or words)? The former 'common' clothing could have less of an impact on identification accuracy due to the increased likelihood that others could be wearing this type of clothing (and thus that the police could easily find an innocent person matching this description). The latter 'distinct' clothing is unlikely to be as typical and thus, its biasing impact may be greater because of the low probability of finding an innocent person who matches this distinct description. If a witness was to engage in the reasoning described above (e.g. 'What are the chances that the police could find an *innocent* person matching my entire description?'), it is likely that he or she would use the distinct clothing as evidence of suspect guilt. Alternatively, common clothing may be less likely to draw the attention of witnesses, less likely to be encoded, and thus less likely to produce any biasing effect. The clothing used in previous clothing bias investigations has clearly been distinctive. Yarmey et al. (1996) used a university T-shirt or sweater (depending on the weather) with a 'Guelph XXL Gryphons' logo on the front (A. D. Yarmey, personal communication, April 30, 2001); Lindsay et al. (1987) used a university sweatshirt with a distinctive logo; Lindsay, et al. (1991) used a T-shirt with stick figures and writing; and Lindsay et al. (1994) used a colourful Hawaiian patterned shirt to demonstrate clothing bias in a mug shot search. Thus, one issue that has not been addressed in the previous literature is the effect of less distinctive clothing on identification bias.

Just as the similarity of the clothing is an issue of some importance, so is the match between the true perpetrator and the innocent suspect. The majority of studies conducted to date have employed as innocent suspects people selected because they resembled the confederates who staged the events. In real cases, the descriptions provided by witnesses

are often vague and, indeed, the description of the perpetrator's clothing may be more detailed and distinctive than the description of the person. As a result, when police apprehend an innocent suspect based on these descriptions, the clothing may match the description of the clothes worn by the perpetrator more closely than the suspect resembles the perpetrator. An important question not addressed in previous studies is the impact of clothing bias on identification of innocent people who are reasonable suspects (i.e. match the general description of the perpetrator) but not particularly similar to the true culprit. If an innocent suspect must closely resemble the perpetrator as well as being dressed similarly for clothing bias to present a problem, the laws of probability suggest that clothing bias may be a real but relatively rare problem for the criminal justice system. On the other hand, if clothing bias produces mistaken identification of people who would not be likely to be identified in the absence of the clothing cues, clothing biased show-ups are a serious threat to innocent suspects and false identifications are likely to occur frequently. The fact that show-ups are recommended in official, government sponsored guidelines for police practice (Technical Working Group for Eyewitness Evidence, 1999) and frequently used by police (Flowe, Ebbesen, Burke, & Chivabunditt, 2001; Gonzalez et al., 1993; McQuiston & Malpass, 2001) make it imperative that the effects of clothing bias on show-ups be investigated further.

Finally, there is no theory concerning the impact of clothing similarity and suspect similarity on the accuracy of show-up identification. One explanation is that clothing provides a memory cue to witnesses assisting them in the search for the face of the criminal and misleading them when the suspect is innocent (literally, a form of guilt by association). This seems unlikely given the consistent reports from lineup studies that correct identification rates are not increased by clothing cues. A second possibility was illustrated above. Clothing information may be used by witnesses in an inferential process where they estimate the likelihood that the person presented is the criminal. A third possibility is that clothing acts as a piece of information in an inferential process but only if the memory trace is weak. Although the current research was not designed to test theoretical explanations for the clothing bias, it is worthwhile to begin discussing the issue.

To explore the issues raised thus far, participants in the current study were exposed to a target person in a field setting. Then, following a delay period, participants were shown either a photograph of the target, a high similarity innocent suspect, or a low similarity (but still plausible) innocent suspect.

METHOD

Participants

A male confederate (hereafter target) approached store and business employees in various shopping malls in southern Ontario, Canada. Only employees who were not occupied (e.g. serving customers or stocking shelves) were approached and asked if they would volunteer to participate in a study for the Department of Psychology at Queen's University. Of the 430 people who agreed to participate (approximately 90% of all persons approached), there were 99 men and 331 women. Some participants (N = 51) were dropped from the analysis because they were no longer available (e.g. gone home for the day or in a meeting) when the experimenter returned to conduct the identification phase of the study.

The final sample (N = 379) consisted of 88 men (*M* age = 28.1 years) and 291 women (*M* age = 27.9 years).

Design

The study employed a 3 (show-up individual: target, high similarity innocent suspect, low similarity innocent suspect) × 2 (type of clothing worn at original event: distinct, common) × 2 (clothing bias: same clothing worn at event and show-up, different clothing at original event and show-up) between-subjects design with time delay as a continuous independent variable. Two additional conditions were used to test a specific issue. The two innocent suspects also were shown in clothing similar, but not identical, to the distinct clothing to investigate the effects of showing innocent people in similar, yet incorrect, clothing. These two conditions will be referred to as 'similar-distinct' conditions and these data were only included in the 'Similar-distinct clothing' section in the Results. The primary dependent measure was identification accuracy. Post-decision confidence also was measured and some description data were obtained.

Materials

Materials were 7.6 × 10.2 cm frontal, head and chest, colour photographs of the target and the high and the low similarity innocent suspects. The target was white, in his early 30s, 1.83 m tall, weighed 113 kg, was clean-shaven, had short brown hair, and no distinguishing or outstanding features. The two innocent suspects were selected based on pilot work where 60 participants interacted with the target for approximately 45 seconds, gave a description of the target, and then ranked from most to least similar the similarity to the target of five foils used in previous lineup research (Pryke, 2001). The foils that were ranked first (i.e. most similar) and fourth overall were selected as the high similarity and low similarity innocent suspects, respectively (Dupuis, 2001). The pilot study results also showed that a few participants ranked the low similarity innocent suspect as most similar to the target, indicating that, although he was dissimilar to the target, he was not an implausible replacement. On the other hand, in lineup research, the low similarity suspect had never been selected from either the target-present or target-absent lineup.

Two shirts were used for the primary clothing bias manipulations. One shirt was a man's blue-gray, short sleeved, button-up plaid shirt (hereafter common clothing). This style of shirt was typical of men's fashion at the time the data were collected and was considered to be common men's apparel. The second shirt was a black Harley-Davidson T-shirt with a motorcycle, blue eagle wings and the Harley-Davidson logo on the front (hereafter distinct clothing). While recruiting participants for the study, the target either wore the common clothing or the distinct clothing. A third shirt was used to specifically test the possibility that similar but not identical clothing may be sufficient to produce a clothing bias effect. The third shirt was a black Harley-Davidson T-shirt with a motorcycle, howling wolf, and the Harley-Davidson logo on the front (hereafter similar-distinct clothing). The size of the design on both Harley-Davidson T-shirts was approximately equal. Photographs of the innocent suspects were taken in each of the three shirts to be used in the identification phase of the study. Photographs of the target in the common and the distinct shirts worn at the event also were used. The two innocent suspects appeared in one of the three shirts at test.

Procedure

Employees (i.e. prospective participants) of stores in shopping malls in southern Ontario, Canada were approached by the target, who introduced himself as a graduate student in psychology. He then asked if they would participate in an experiment for the Department of Psychology at Queen's University. Employees (in groups of one to three people) were told that, if they agreed to participate, an (female) experimenter would return later that day to ask them a few innocuous questions. Because we wanted to avoid telling participants that the study was investigating their ability to identify the target (in case this type of knowledge affected the participants' accuracy) and we needed to give participants enough information to decide if they wanted to participate in the study, the target explained that he could not tell participants what the questions were because his job was only to recruit participants. However, the target assured employees that no personal or embarrassing questions would be asked of them and that no questions would be asked regarding their place of employment. At this point, if store employees agreed to take part in the study, the target thanked them for their time and reminded them that a second person would return later that day. If employees did not want to participate or felt that they needed permission from management prior to participating in the study, they were told of the purpose of the study and were not included as participants. Immediately upon leaving each store where employees had agreed to participate, the target recorded the time, the name of the store, and the sex of the participant, approximate age, and a brief description of the participant to facilitate the identification of participants by the returning experimenter. This was an important step because not all store employees who were present when the experimenter returned were present when the target had earlier recruited participants.

The experimenter entered the stores and recorded the time which, when compared with the confederate's recorded time, indicated delays varying from 10 to 272 minutes (to the nearest minute). The experimenter assessed whether the participant(s) were unoccupied and, therefore, available to complete the study. At this time, if participants were occupied, the experimenter returned later or waited until the participant(s) were available (recording that time). If the participant(s) were unoccupied, they were told that they were going to be shown a photograph and asked to make a decision as to whether or not the photograph was of the person who asked them to participate in the study. Participants were told that the photograph they were going to be shown may or may not be of the person who recruited them to be in the study (standard lineup warning). A photograph of the target, the high similarity innocent suspect, or the low similarity innocent suspect was then shown to the participant. The person in the photograph was either wearing the common, distinct, or similar-distinct (innocent suspects only) clothing. Participants then made an identification decision and rated their confidence in that decision on a 10-point scale (1 being 'Not at all confident', 10 being 'Extremely confident'). If customers/ clients had entered the store during the identification and participants needed to be dismissed, the study ended immediately after obtaining their confidence rating. However, if no customers awaited service, participants were asked, while viewing the photograph, if they could remember and describe what the target was wearing when he came into the store.

Participants were randomly assigned to conditions where possible. It was logistically impossible for the confederate to change shirts from one store to the next. As a result, the confederate would recruit participants wearing either the common or the distinct shirt for a

J. E. Dysart et al.

period of time and then change to the other shirt. Within groups of participants recruited during each block of time, show-up conditions were assigned randomly. Differences in delay were haphazard rather than random but mean delay was approximately equal (not significantly different) across clothing and show-up conditions (all F values <1).

RESULTS

The results presented below tested a 3 (show-up individual) $\times$ 2 (clothing bias) $\times$ 2 (clothing worn at event) between-subjects design with identification accuracy from show-ups as the primary dependent variable. The data from two additional conditions, where the innocent suspects were presented in similar looking Harley-Davidson (distinct) attire were not included in these analyses and are presented separately in the section titled 'Similar-distinct clothing'.

Identification accuracy

Delay
Logistic regression analyses revealed no effects of delay on identification accuracy from show-ups or any significant interactions with delay in any condition. Participants were equally likely to make a correct identification decision from 10 minutes to 4 hours after the event. Therefore, the remainder of the analyses were conducted by collapsing over this variable.

Overall
Using logistic regression analysis with accuracy as the dependent variable and all of the variables in the equation, a three-way interaction was tested between clothing bias, type of clothing at event and photograph shown. Results showed no significant interaction ($Wald = 1.68$, $p = 0.195$) of these variables. Overall, there was a significant main effect of bias on identification accuracy, with the biased conditions yielding a lower correct decision rate (74%) than the unbiased conditions (87%; $Z = 3.22$, $p = 0.001$). However, logistic regression analysis, with clothing type and clothing bias in the equation, indicated a significant clothing type by clothing bias interaction ($Wald = 10.29$, $p = 0.001$). Participants in the distinct clothing condition were significantly less likely to make a correct identification decision from the biased (64%) than from the unbiased (90%) condition ($Z = 4.03$, $p < 0.001$), whereas participants in the common clothing condition were as likely to make a correct identification decision from the biased (89%) and unbiased conditions (84%; $Z = 0.96$, $p = 0.34$).

Additional interactions were tested by separating the target-present and two target-absent conditions. The clothing bias by type of clothing interaction (with both variables in the equation), for the target-present and low similarity innocent suspect were non-significant ($Wald = 2.72$, $p = 0.10$, $Wald = 0.01$, $p = 0.99$, respectively). The result of the same analysis with the high similarity innocent suspect, however, was significant ($Z = 3.71$, $p = 0.05$). Participants in the unbiased conditions were more accurate in the distinct clothing condition (92.6%) over the common clothing condition (84.0%), whereas participants in the biased conditions were more accurate in the common clothing condition (96.0%) than in the distinct clothing condition (83.3%).

Table 1. Proportion of 'Yes' identification decisions for common and distinct clothing as a function of clothing bias conditions and show-up individual

Clothing at event	Clothing bias condition		
	Unbiased	Biased	Biased similar-distinct
Common clothing			
Show-up individual			
Target	0.84 (21/25)	0.96 (24/25)	N/A
High similarity innocent suspect	0.25 (6/24)	0.16 (4/25)	N/A
Low similarity innocent suspect	0.08 (2/25)	0.13 (3/24)	N/A
Distinct clothing			
Show-up individual			
Target	0.93 (25/27)	0.83 (20/24)	N/A
High similarity innocent suspect	0.23 (7/30)	0.50 (17/34)	0.50 (14/28)
Low similarity innocent suspect	0.00 (0/29)	0.37 (11/30)	0.14 (4/29)

Note: Frequencies shown in parentheses.

Distinct clothing

A significant bias by photograph interaction was found within the distinct clothing conditions ($Wald = 16.64, p < 0.001$). Correct identifications of the target were unaffected by the clothing bias manipulation in the distinct clothing conditions, with 83% of participants identifying him in the biased condition and 93% identifying him from the unbiased condition (see Table 1; $Z = 1.01, p = 0.31$). On the other hand, identification accuracy for the high similarity and low similarity innocent suspects was significantly affected by the clothing bias manipulations. When the high similarity innocent suspect was presented in the unbiased condition, 23% of participants incorrectly identified him, whereas 50% of participants identified him from the biased condition ($Z = 2.21, p = 0.01$). Similar results were obtained for the low similarity innocent suspect, where 0% and 37% of participants identified him from the unbiased and biased conditions, respectively ($Z = 3.61, p = 0.001$).

Common clothing

No interaction was found between the clothing bias conditions and photograph shown ($Wald = 0.15, p = 0.70$) when the target appeared in the common clothing. Participants who viewed the biased target-present show-up were no more accurate (96%) than those who viewed the unbiased target-present show-up (84%; $Z = 1.41, p = 0.16$; see Table 1). Participants who were shown a photograph of the high similarity innocent suspect were unaffected by the biased clothing manipulation, with 16% and 25% identifying the innocent suspect in the biased and unbiased conditions, respectively ($Z = 0.78, p = 0.44$). Finally, participants who viewed the low similarity innocent suspect were also unaffected by innocent suspect attire, with 13% and 8% of participants identifying the innocent suspect from the biased and unbiased conditions, respectively ($Z = 0.52, p = 0.60$).

Similar-distinct clothing

Two additional conditions were run in which the high and low similarity innocent suspects were presented in a Harley-Davidson T-shirt that was similar to, but not the same as, the T-shirt worn by the target. As explained earlier, this was done to explore the effects of

presenting an innocent person in clothing that matches the description provided by the witness to some extent, but is not the exact clothing worn by the target. When the high similarity innocent suspect was presented in the similar-distinct clothing, 50% of participants identified him as the target. This proportion of identifications was not significantly different from that obtained in the biased (i.e. same Harley-Davidson) T-shirt condition (50%; $Z = 0.00$, $p = 1.00$). When the low similarity innocent suspect was presented in the similar-distinct clothing, 14% of participants identified him as the target, a significantly greater percentage than those who viewed the unbiased photograph (0%; $Z = 2.07$, $p = 0.02$), and a significantly lower percentage than those who viewed him in the biased photograph condition (37%; $Z = 2.02$, $p = 0.02$).

Clothing descriptions

The primary issue of interest for the accuracy of clothing descriptions is whether participants who gave a correct description of the target's clothing were more likely to make a correct identification decision. It should be noted again, however, that not all participants provided a description of the target. Participants who clearly needed to be excused from the study after the primary dependent measures had been collected (e.g. customers had entered the store during data collection and needed assistance) were dismissed after their confidence judgement had been given. In total, 66.2% of participants in the distinct conditions (N = 153) and 48.6% of participants in the common clothing conditions (n = 72) provided a description of the target's clothing.

Common clothing
A description of the common clothing was considered correct if the participant said that the target wore a blue/grey button-up or plaid shirt. Overall, 52.7% of participants correctly described the common clothing. An additional 31.9% of participants gave an incorrect description (e.g. light T-shirt), 6.9% gave an insufficient description (e.g. dark shirt), and 8.3% stated that they could not remember what the target was wearing. Identification accuracy was unaffected by whether or not a participant gave a correct (89%) or incorrect (78%) clothing description ($Z = 1.11$, $p = 0.13$).

Distinct clothing
A description of the distinct clothing was considered correct if the participant stated that the target wore a black T-shirt. Overall, 62% of participants correctly described the distinct clothing, 28% of participants gave an incorrect description (e.g. red T-shirt), 4% gave a correct but very limited description (e.g. T-shirt), and 7% stated that they could not remember what the target was wearing. Participants who provided a correct description of the distinct clothing were no more likely to make a correct identification decision (83%) than were participants who gave an incorrect description (74%; $Z = 1.17$, $p = 0.12$).

Confidence

To examine the relationship between eyewitness confidence and accuracy, we distinguished between eyewitnesses who chose the suspect from the showup (choosers) and those who did not choose the suspect from the showup (non-choosers). For choosers, accurate eyewitnesses were more confident ($M = 8.88$) than inaccurate eyewitnesses ($M = 7.87$; $F(1, 156) = 21.28$, $p < 0.001$). The confidence-accuracy correlation was

0.32. Similarly, accurate non-choosers were more confident ($M = 7.88$) than inaccurate non-choosers ($M = 5.96$; $F(1, 219) = 12.29$, $p = 0.001$). The confidence-accuracy correlation was 0.19. Breaking the conditions down further (biased vs. unbiased showups, distinctive vs. common clothing, etc.) produced correlations ranging from 0.09 to 0.42. The size and range of the obtained correlations is consistent with the results of lineup research.

DISCUSSION AND CONCLUSIONS

Several interesting results were obtained in the current study. First, there was a significant difference in overall identification accuracy from common and distinct clothing bias manipulations. When the target wore what was considered to be common clothing, the effects of clothing bias were non-existent. However, when the target wore clothing that was considered to be distinct, the effect of clothing bias on identification accuracy was profound. A second intriguing finding from the current study pertained to the similar-distinct clothing manipulation, where the clothing worn during the event and that shown in the identification procedure were different but similar. The results from these conditions suggest that if a person who resembles the perpetrator is apprehended near the scene of the crime, and is wearing distinct clothing similar to that described by the eyewitness, the likelihood of false identification is considerable. Furthermore, these false identifications are as likely whether an innocent person is wearing identical distinct clothing to that worn by the perpetrator or merely similar clothing. A similar pattern of results was found for the dissimilar, yet plausible looking innocent suspect. In this situation, the false identification rate with the similar-distinct clothing was significantly higher than that obtained in the non-biased condition. Together, these results suggest that the effect of distinct clothing bias on identification accuracy from show-ups is substantial. The fact that the dissimilar innocent suspect was never identified unless he was wearing similar or identical clothing is a clear indication of how powerful the clothing bias can be in identification from show-ups.

Future research needs to be done which addresses issues not yet raised in this paper, including the ability of witnesses to identify clothing from other similar looking items. For example, would it be possible for a witness to distinguish between a lineup of white T-shirts? Or short-sleeved plaid dress-shirts? This could be tested for in any investigation using clothing lineups to help determine if the item is unique enough to be distinguishable from other similar items.

Another important issue that remains unexplored is whether a clothing bias will exist when the witness' description of the clothing is incorrect. In all bias research to date, clothing bias has been manipulated by presenting innocent suspects in attire that matches the clothing **actually** worn during the event (either an exact match or highly similar). When dissimilar clothing is used (control conditions), it is selected to differ considerably from the attire worn during the event. In real cases, police must rely on the witness' description of clothing. As the results in this paper indicate, witnesses are not perfect in their descriptions of clothing. In real world cases this may lead to occasions when police detain a suspect because his clothing matches an inaccurate description provided by an eyewitness. Only further research will determine if the clothing bias effect is present in these situations.

The salience of the clothing worn during the event also raises several interesting issues. Context may influence how salient clothing is to the witness. Strikingly salient clothing in one context (e.g. a military uniform on a university campus) may not be salient at all in another context (e.g. on a military base). This may present serious problems for police and

J. E. Dysart et al.

courts, as it may be impossible to determine if clothing bias is likely to occur in a given case because it will be difficult to decide if the clothing worn by the criminal was distinctive. As well, salience may vary with witness characteristics. Using the previous example, a military uniform may be either more or less salient to a member of the military than to a civilian.

As with all research, there are limitations that must be acknowledged. The current study relied on a single confederate and thus lacks stimulus generalizability (Wells & Windschitl, 1999). Identification procedures may interact with the appearance of confederates and criminals in non-obvious ways. Our confederate was a large man. His size may have influenced the witnesses either as a memory cue or via conscious inferences. Thus, the fact that such a person is less common than men of slighter build may have increased the impact of the clothing bias. Replication and extension of the findings with other confederates will be necessary to alleviate this concern. Similarly, the high and low similarity innocent suspects were specific individuals and the results may have differed if others had taken their place. Again, only further research can determine how important the degree of similarity to the criminal is in this situation.

In addition, the identification procedure in the current study used photographs to present the targets and innocent suspects, whereas, Flowe et al.'s (2001) results showed that the majority of real-world show-up identifications were conducted live. Live presentation may assist witnesses, helping them to realize that the clothing is not identical to the clothing they saw and thus leading to an inference of innocence. Alternatively, live show-ups may produce a more powerful effect if the reasoning process described earlier was being used by real witnesses. If witnesses ask themselves 'What are the chances that the police could find someone so quickly who matched my description?', they may conclude that the probability is lower when the person is presented live leading to a stronger inference of guilt.

Another limitation of the current study is that the exposure to the target did not involve a 'criminal act', like the majority of laboratory studies that have been conducted over the past 20 years. However, the live, staged-crime thefts that are conducted in eyewitness laboratories cannot be conducted safely in field settings. The likelihood that witnesses in field settings would phone the police or alert security after a 'theft', or even chase after a 'thief' is great, and thus not a wise practice. It also could be argued that witnesses in laboratory studies are usually made aware immediately after a staged event that no real crime has occurred. Thus, their interactions with the target person are effectively no different from the field setting witnesses, with the exception that there is a speedy exit of the target person in laboratory studies. Although there may be a small trade-off between using laboratory staged crimes and field studies, the exposure to the target is, for all practical purposes, similar. In addition, not all crimes that occur in the real-world involve a quick exit of the criminal or knowledge on the part of the witness that a crime has actually occurred. For example, victims of fraud are not aware that their interaction with the person committing fraud is actually a crime. In fact, they may only come to realize this weeks or months after the event has occurred. This type of crime is very similar to the situation used in this study.

Live crimes also can be actually or potentially violent and thus arousing or threatening to the witness. Very few studies mimic this aspect of real crimes. On the other hand, the available data do not support the conclusion that eyewitness accuracy is enhanced by threat or violence (Steblay, 1992). Even potentially violent crimes can produce witnesses who were neither threatened nor aroused. For example, a few years ago a colleague opened a

door for a man who was limping. Only later, after the man had left the scene, did he learn that the man was limping because he had a shotgun concealed in his pants. He had used the shotgun to rob the bank they had entered together. Although low arousal identification situations are very common and worthy of study and potentially arousing crimes may not generate arousal in all witnesses, the fact remains that the results of this research need not generalize to witnesses who are aroused.

Finally, the distinction between common and distinct clothing is ill defined, relying exclusively on the judgement of the authors. To the extent that this reflects a concern about the salience of the clothing within the context of the event and not just a property of the clothing itself, further research on this issue may be difficult. Salience has a tendency to be defined circularly; stimuli that are salient are those that have an effect and stimuli that do not have an effect were not salient (e.g. Taylor & Fiske, 1978). On the other hand, the different shirts did result in different patterns of correct and incorrect identification decisions and the pattern of effects reported seem to be interpretable in terms of salience or distinctiveness.

In real world cases it will be difficult or impossible to be certain that a clothing description is accurate and to know if the clothing was salient to the witness (though spontaneously providing a clothing description would suggest that it was). The results from the current study suggest that clothing bias can have a powerful impact on the accuracy of show-up identification decisions. Although future studies should investigate the most effective means of avoiding clothing bias, this problem could potentially be avoided by not exposing the witness to the suspect's clothing during an identification procedure. One way of doing this may be to cover the suspect from the neck down with a blanket or jacket. If a photo is used, the picture could be restricted to the face only or by covering the clothing before showing the photo to the witness. However, these suggestions are tentative, and one must be cautious in conducting such a procedure prior to empirical investigations being conducted which show that covering clothing information, and thus effectively covering body cues (e.g. stature, weight), does not reduce identification accuracy. Until a satisfactory procedure is found and successfully tested, any showup procedure conducted where clothing cues are visible may potentially fall prey to the clothing bias effect.

ACKNOWLEDGEMENTS

REFERENCES

Brehm, S., & Brehm, J. W. (1981). *Psychological reactance: A theory of freedom and control.* New York: Academic Press.

Clark, S. E., & Davey, S. L. (2005). The target-to-foils shift in simultaneous and sequential lineups. *Law and Human Behavior, 29,* 151–172.

Dekle, D. J., Beal, C. R., Elliott, R., & Huneycutt, D. (1996). Children as witnesses: A comparison of lineup versus show-up identification methods. *Applied Cognitive Psychology, 10,* 1–12.

Dupuis, P. R. (2001). Calibrating eyewitness identification accuracy using rank-combined lineups. Unpublished Master's thesis, Queen's University, Kingston, Ontario, Canada.

Dysart, J. E., & Lindsay, R. C. L. (in press). Suggestive technique or reliable method: What do we really know about show-up identifications? In R. C. L. Lindsay, D. F. Ross, J. D. Read, & M. Toglia (Eds.), *Handbook of eyewitness psychology: Memory for people*.

Flowe, H., Ebbesen, E., Burke, C., & Chivabunditt, P. (2001 June). At the scene of the crime: An examination of the external validity of published studies on lineup identification accuracy. *Paper Presented at the Annual Meeting of the American-Psychological Society*, Toronto, Ontario, Canada.

Flynn, K. (1998, October 23). Persevering woman helps free stranger in '90 murder case. *New York Times*, p. 1.

Gonzalez, R., Ellsworth, P. C., & Pembroke, M. (1993). Response biases in lineups and show-ups. *Journal of Personality and Social Psychology, 64*, 525–537.

Lindsay, R. C. L., & Bellinger, K. (1999). Alternatives to the sequential lineup: The importance of controlling the pictures. *Journal of Applied Psychology, 84*, 315–321.

Lindsay, R. C. L., & Wells, G. L. (1980). What price justice? Exploring the relationship of lineup fairness to identification accuracy. *Law and Human Behavior, 4*, 303–313.

Lindsay, R. C. L., & Wells, G. L. (1985). Improving eyewitness identification from lineups: Simultaneous versus sequential lineup presentation. *Journal of Applied Psychology, 70*, 556–564.

Lindsay, R. C. L., Lea, J., & Fulford, J. (1991). Sequential lineup presentation: Technique matters. *Journal of Applied Psychology, 76*, 741–745.

Lindsay, R. C. L., Wallbridge, H., & Drennan, D. (1987). Do the clothes make the man?: An exploration of the effect of lineup attire on eyewitness identification accuracy. *Canadian Journal of Behavioral Science, Special Issue: Forensic Psychology, 19*, 463–478.

Lindsay, R. C. L., Nosworthy, G. J., Martin, R., & Martynuck, C. (1994). Using mug shots to find suspects. *Journal of Applied Psychology, 79*, 121–130.

Lindsay, R. C. L., Pozzulo, J. D., Craig, W., Lee, K., & Corber, S. (1997). Simultaneous lineups, sequential lineups, and show-ups: Eyewitness identification decisions of adults and children. *Law and Human Behavior, 21*, 391–404.

Lindsay, R. C. L., Lea, J. A., Nosworthy, G. J., Fulford, J. A., Hector, J., LeVan, V., & Seabrook, C. (1991). Biased lineups: Sequential presentation reduces the problem. *Journal of Applied Psychology, 76*, 796–802.

Malpass, R. S., & Devine, P. G. (1981). Eyewitness identification: Lineup instructions and the absence of the offender. *Journal of Applied Psychology, 66*, 482–489.

McQuiston, D. E., & Malpass, R. S. (2001). An archival review of eyewitness identifications in criminal felony cases. *Paper Presented at the 4th Biennial Conference for the Society for Applied Research in Memory and Cognition*, Kingston, Ontario, Canada.

Phillips, M. R., McAuliff, B. D., Kovera, M. B., & Cutler, B. L. (1999). Double-blind photoarray administration as a safeguard against investigator bias. *Journal of Applied Psychology, 84*, 940–951.

Pryke, S. (2001). Multiple independent identification decisions: A method of calibrating eyewitness identifications. Doctoral dissertation, Queen's University, Kingston, Ontario, Canada.

Steblay, N. M. (1992). A meta-analytic review of the weapon focus effect. *Law and Human Behavior, 16*, 413–424.

Steblay, N., Dysart, J. E., Fulero, S., & Lindsay, R. C. L. (2001). Eyewitness accuracy rates in sequential and simultaneous lineup presentations: A meta-analytic comparison. *Law and Human Behavior, 25*, 459–473.

Steblay, N., Dysart, J. E., Fulero, S., & Lindsay, R. C. L. (2003). Eyewitness accuracy rates in police showup and lineup presentations: A meta-analytic comparison. *Law and Human Behavior, 27*, 523–540.

Taylor, S. E., & Fiske, S. T. (1978). Salience, attention, and attributions: Top of the head phenomena. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 11). New York: Academic Press.

Technical Working Group for Eyewitness Evidence. (1999). *Eyewitness evidence: A guide for law enforcement*. Washington, DC: U.S. Department of Justice, Office of Justice Programs.

Wagenaar, W. A., & Veefkind, N. (1992). Comparison of one-person and many-person lineups: A warning against unsafe practices. In F. Losel, D. Bender, & T. Bliesener (Eds.), *Psychology and law: International perspectives* (pp. 275–285). Berlin: de Gruyter.

Wells, G. L. (1984). The psychology of lineup identification. *Journal of Applied Social Psychology, 14*, 89–103.

Wells, G. L., & Windschitl, P. D. (1999). Stimulus sampling and social psychological experimentation. *Personality and Social Psychology Bulletin, 25*, 1115–1125.

Yarmey, A. D., Yarmey, A. L., & Yarmey, M. J. (1994). Face and voice identifications in show-ups and lineups. *Applied Cognitive Psychology, 8*, 453–464.

Yarmey, A. D., Yarmey, M. J., & Yarmey, A. L. (1996). Accuracy of eyewitness identifications in show-ups and lineups. *Law and Human Behavior, 20*, 459–477.

Leading Questions and the Eyewitness Report

Elizabeth F. Loftus[1]
University of Washington

A total of 490 subjects, in four experiments, saw films of complex, fast-moving events, such as automobile accidents or classroom disruptions. The purpose of these experiments was to investigate how the wording of questions asked immediately after an event may influence responses to questions asked considerably later. It is shown that when the initial question contains either true presuppositions (e.g., it postulates the existence of an object that did exist in the scene) or false presuppositions (e.g.. postulates the existence of an object that did not exist), the likelihood is increased that subjects will later report having seen the presupposed object. The results suggest that questions asked immediately after an event can introduce new — not necessarily correct — information, which is then added to the memorial representation of the event, thereby causing its reconstruction or alteration.

1   Although current theories of memory are derived largely from experiments involving lists of words or sentences, many memories occurring in everyday life involve complex, largely visual, and often fast-moving events. Of course, we are rarely required to provide precise recall of such experiences — though as we age, we often volunteer them — but on occasion such recall is demanded, as when we have witnessed a crime or an accident. Our theories should he able to encompass such socially important forms of memory. It is clearly of concern to the law, to police and insurance investigators, and to others to know something about the completeness, accuracy, and malleability of such memories.

2   When one has witnessed an important event, one is sometimes asked a series of questions about it. Do these questions, if asked immediately after the event, influence the memory of it that then develops? This paper first summarizes research suggesting that the wording of such initial questions can have a substantial effect on the answers given, and then reports four new studies showing that the wording of these initial questions can also influence the answers to different questions asked at [begin page 561] some later time. The discussion of these findings develops the thesis that questions asked about an event shortly after it occurs may distort the witness' memory for that event.

Answers Depend on the Wording of Questions

3   An example of how the wording of a question can affect a person's answer to it has been reported by Harris (1973). His subjects were told that "the experiment was a study in the accuracy of guessing measurements, and that they should make as intelligent a numerical guess as possible to each question" (p. 399). They were then asked either of two questions such as, "How tall was the basketball player?", or, "How short was the basketball player?" Presumably the former form of the question presupposes nothing about the height of the player, whereas

the latter form involves a presupposition that the player is short. On the average, subjects guessed about 79 and 69 in. (190 and 175 mm), respectively. Similar results appeared with other pairs of questions. For example, "How long was the movie?", led to an average estimate of 130 min, whereas, "How short was the movie?" led to 100 min. While it was not Harris' central concern, his study clearly demonstrates that the wording of a question may affect the answer.

4   The phenomenon has also been demonstrated in two other contexts: past personal experiences and recently-witnessed events.

Past Personal Experiences

5   In one study (Loftus, unpublished), 40 people were interviewed about their headaches and about headache products under the belief that they were participating in market research on these products. Two of the questions were crucial to the experiment. One asked about products other than that currently being used, in one of two wordings:

(1a) In terms of the total number of products, how many other products have you tried? 1? 2? 3?

(1b) In terms of the total number of products, how many other products have you tried? 1? 5? 10?

6   The 1/2/3 subjects claimed to have tried an average of 3.3 other products, whereas the 1/5/10 subjects claimed an average of 5.2; $t(38) = 3.14$, $\sigma = .61$, $p < .01$.

7   The second key question asked about frequency of headaches in one of two ways:

(2a) Do you get headaches frequently, and, if so, how often?

(2b) Do you get headaches occasionally, and, if so, how often?

8   The "frequently" subjects reported an average of 2.2 headaches/wk, whereas the "occasionally" group reported only 0.7/wk; $t(38) = 3.19$, $\sigma = .47$, $p < .01$. [begin page 562]

Recently Witnessed Events

9   Two examples from the published literature also indicate that the wording of a question put to a person about a recently-witnessed event can affect a person's answer to that question. In one study (Loftus, 1974; Loftus & Zanni, 1975), 100 students viewed a short film segment depicting a multiple-car accident. Immediately afterward, they filled out a 22-item questionnaire which contained six critical questions. Three of these asked about items that had appeared in the film whereas the other three asked about items not present in the film. For half the subjects, all the critical questions began with the words, "Did you see a . . ." as in, "Did you see a broken headlight?" For the remaining half, the critical questions began with the words, "Did you see the . . ." as in, "Did you see the broken headlight?"

10   Thus, the questions differed only in the form of the article, *the* or *a*. One uses "the" when one assumes the object referred to exists and may be familiar to the listener. An investigator who asks, "Did you see the broken headlight?" essentially says, "There was a broken headlight. Did you happen to see it?" His assumption may influence a witness' report. By contrast, the article "a" does not necessarily convey the implication of existence.

11   The results showed that witnesses who were asked "the" questions were more likely to report having seen something, whether or not it had really appeared in the film, than those who were asked "a" questions. Even this very subtle change in wording influences a witness' report.

12   In another study (Loftus & Palmer, 1974), subjects saw films of automobile accidents and then answered questions about the accidents. The wording of a question was shown to affect a numerical estimate. In particular, the question, "About how fast were the cars going when they smashed into each other?" consistently elicited a higher estimate of speed than when "smashed" was replaced by "collided," "bumped," "contacted," or "hit."

13   We may conclude that in a variety of situations the wording of a question about an event can influence the answer that is given. This effect has been observed when a person reports about his own experiences, about events he has

recently witnessed, and when answering a general question (e.g., "How short was the movie?") not based on any specific witnessed incident.

Question Wording and Answers to Subsequent Questions

14 Our concern in this paper is not on the effect of the wording of a question on its answer, but rather on the answers to other questions asked some time afterward. We will interpret the evidence to be presented as suggesting a memorial phenomenon of some importance. [begin page 563]

15 In the present experiments, a key initial question contains a *presupposition*, which is simply a condition that must hold in order for the question to be contextually appropriate. For example, the question, "How fast was the car going when it ran the stop sign?" presupposes that there was a stop sign. If a stop sign actually did exist, then in answering this question a subject might review, strengthen, or make more available certain memory representations corresponding to the stop sign. This being the case, the initial question might be expected to influence the answer to a subsequent question about the stop sign, such as the question, "Did you see the stop sign?" A simple extension of the argument of Clark and Haviland (in press) can be made here: When confronted with the initial question, "How fast was the car going when it ran the stop sign?", the subject might treat the presupposed information as if it were an address, a pointer, or an instruction specifying where information related to that presupposition may be found (as well as where new information is to be integrated into the previous knowledge). In the process the presupposed information may be strengthened.

16 What if the presupposition is false? In that case it will not correspond to any existing representation, and the subject may treat it as new information and enter it into his memory. Subsequently, the new "false" information may appear in verbal reports solicited from the subject.

17 To explore these ideas, subjects viewed films of complex, fast-moving events. Viewing of the film was followed by initial questions which contained presuppositions that were either true (Experiment 1) or false (Experiments 2-4). In Experiment I, the initial questions either did or did not mention an object that was in fact present in the film. A subsequent question, asked a few minutes later, inquired as to whether the subject has seen the existing object. In Experiments 2-4, the initial questions were again asked immediately after the film, whereas the subsequent questions were asked after a lapse of 1 wk.

Experiment 1

Method

18 One hundred and fifty University of Washington students, in groups of various sizes, were shown a film of a multiple-car accident in which one car, after failing to stop at a stop sign, makes a right-hand turn into the main stream of traffic. In an attempt to avoid a collision, the cars in the oncoming traffic stop suddenly and a five-car, bumper-to-bumper collision results. The film lasts less than 1 min, and the accident occurs within a 4-sec period.

19 At the end of the film, a 10-item questionnaire was administered. A diagram of the situation labeled the car that ran the stop sign as "A," and the cars involved in the collision as "B" through "F." The first [begin page 564] question asked about the speed of Car A in one of two ways:

1. How fast was Car A going when it ran the stop sign?
2. How fast was Car A going when it turned right?

20 Seventy-five subjects received the "stop sign" question and 75 received the "turned right" question. The last question was identical for all subjects: "Did you see a stop sign for Car A?" Subjects responded by circling "yes" or "no" on their questionnaires.

Results and Discussion

21 Fifty-three percent of the subjects in the "stop sign" group responded "yes" to the question, "Did you see a stop sign for Car A?", whereas only 35% in the "turn right" group claimed to have seen the stop sign; χ^2 (1) = 4.98, p < .05. The wording of a presupposition into a question about an event, asked immediately after that event has taken place, can influence the answer to a subsequent question concerning the presupposition itself, asked a very short time later, in the direction of conforming with the supplied information.

22 There are at least two possible explanations of this effect. The first is that when a subject answers the initial stop sign question, he somehow reviews, or strengthens, or in some sense makes more available certain memory representations corresponding to the stop sign. Later, when asked, "Did you see a stop sign . . . ?", he responds on the basis of the strengthened memorial representation.

23 A second possibility may be called the "construction hypothesis." In answering the initial stop sign question, the subject may "visualize" or "reconstruct" in his mind that portion of the incident needed to answer the question, and so, if he accepts the presupposition, he introduces a stop sign into his visualization whether or not it was in memory. When interrogated later about the existence of the stop sign, he responds on the basis of his earlier supplementation of the actual incident. In other words, the subject may "see" the stop sign that he has himself constructed. This would not tend to happen when the initial question refers only to the right turn.

24 The construction hypothesis has an important consequence. If a piece of true information supplied to the subject after the accident augments his memory, then, in a similar way, it should be possible to introduce into memory something that was not in fact in the scene, by supplying a piece of false information. For example, Loftus and Palmer (1974, Expt. 2) showed subjects a film of an automobile accident and followed it by questions about events that occurred in the film. Some subjects were asked "About how fast were the cars going when they smashed into each other?", whereas others were asked the same question with "hit" substituted for "smashed." On a retest 1 wk later, those questioned with "smashed" were more likely than those questioned with "hit" to agree [begin page 565] that they had seen broken glass in the scene, even though none was present in the film. In the present framework, we assume that the initial representation of the accident the subject has witnessed is modified toward greater severity when the experimenter uses the term "smashed" because the question supplies a piece of new information, namely, that the cars did indeed smash into each other. On hearing the "smashed" question, some subjects may reconstruct the accident, integrating the new information into the existing representation. If so, the result is a representation of an accident in memory that is more severe than, in fact, it actually was. In particular, the more severe accident is more likely to include broken glass.

25 The presupposition that the cars smashed into each other may be additional information, but it can hardly be said to be false information. It is important to determine whether it is also true that false presuppositions can affect a witness' answer to a later question about that presupposition. Such a finding would imply that a false presupposition can be accepted by a witness, that the hypothesis of a strengthening of an existing memorial representation is untenable (since there should be no representation corresponding to nonexistent objects), and that the construction hypothesis discussed above is supported. Experiment 2 was designed to check this idea.

Experiment 2

Method

26 Forty undergraduate students at the University of Washington, again in groups of various sizes, were shown a 3-min videotape taken from the film *Diary of a Student Revolution*. The sequence depicted the disruption of a class by eight demonstrators; the confrontation, which was relatively noisy, resulted in the demonstrators leaving the classroom.

27 At the end of the videotape, the subjects received one of two questionnaires containing one key and nineteen filler questions. Half of the subjects were asked, "Was the leader of the four demonstrators who entered the classroom a male?", whereas the other half were asked, "Was the leader of the twelve demonstrators who entered the

classroom a male?" The subjects responded by circling "yes" or "no."

28 One week later, all subjects returned and, without reviewing the videotape, answered a series of 20 new questions about the disruption. The subjects were urged to answer the questions from memory and not to make inferences. The critical question here was, "How many demonstrators did you see entering the classroom?"

Results and Discussion

29 Subjects who had previously been asked the "12" question reported having seen an average of 8.85 people 1 wk earlier, whereas those asked [begin page 566] the "4" question recalled 6.40 people, t (38) = 2.50, σ = .98, $p < .01$. The actual number was, it will be recalled, eight. One possibility is that some fraction of the subjects remembered the number 12 or the number 4 from the prior questionnaire and were responding to the later question with that number, whereas the remainder had the correct number. An analysis of the actual responses given reveals that 10% of the people who had been interrogated with "12" actually responded "12," and that 10% of those interrogated with "4" actually responded with "4." A recalculation of the means, excluding those subjects in the "12" condition who responded "12" and those in the "4" condition who responded "4," still resulted in a significant difference between the two conditions (8.50 versus 6.67), $t(34) = 1.70$, $p < .05$. This analysis demonstrates that recall of the specific number given in the initial questionnaire is not an adequate alternative explanation of the present results.

30 The result shows that a question containing a false numerical presupposition can, on the average, affect a witness' answer to a subsequent question about that quantitative fact. The next experiment was designed to test whether the same is true for the existence of objects when the false presupposition concerns one that did not actually exist.

31 Experiment 3

Method

32 One hundred and fifty students at the University of Washington, in groups of various sizes, viewed a brief videotape of an automobile accident and then answered ten questions about the accident. The critical one concerned the speed of a white sports car. Half of the subjects were asked, "How fast was the white sports car going when it passed the barn while traveling along the country road?", and half were asked, "How fast was the white sports car going while traveling along the country road?" In fact, no barn appeared in the scene.

33 All of the subjects returned 1 wk later and, without reviewing the videotape, answered ten new questions about the accident. The final one was, "Did you see a barn?" The subjects responded by circling "yes" or "no" on their questionnaires.

Results and Discussion

34 Of the subjects earlier exposed to the question containing the false presupposition of a barn, 17.3% responded "yes" when later asked, "Did you see a barn?", whereas only 2.7% of the remaining subjects claimed to have seen it; χ^2 (1) = 8.96, $p < .01$. An initial question containing a false presupposition can, it appears, influence a witness' later tendency to report the presence of the nonexistent object corresponding to that presupposition. [begin page 567]

35 The last experiment not only extends this finding beyond the single example, but asks whether or not the effect is wholly due to the word "barn" having occurred or not occurred in the earlier session. Suppose an initial question merely asks about, instead of presupposing, a nonexistent object; for example, "Did you see a barn?," when no barn existed. Presumably subjects will mostly respond negatively to such questions. But, what if that same question is asked again some time later? It is possible that a subject will reflect to himself, "I remember something about a barn, so I guess I must have seen one." If this were the case, then merely asking about a nonexistent object could increase the tendency to report the existence of that object at some later time, thereby accounting for the results of Expt III.

Experiment 4

Method

36   One hundred and fifty subjects from the University of Washington, run in groups of various sizes, viewed a 3-min 8 mm film clip taken from inside of an automobile which eventually collides with a baby carriage being pushed by a man. Following presentation of the film, each subject received one of three types of booklets corresponding to the experimental conditions. One hundred subjects received booklets containing five key and 40 filler questions. In the "direct" version, the key questions asked, in a fairly direct manner, about items that were not present in the film. One example was, "Did you see a school bus in the film?" All of these questions are listed in Table I, under the column labeled "Direct questions." In the "False presupposition" version, the key questions contained false presuppositions referring to an item that did not occur in the film. The corresponding example was, "Did you see the children getting on the school bus?" All of these questions are listed in Table I under the column labeled "False presupposition questions." The third group of 50 subjects received only the 40 filler questions and no key questions. The goal of using so many filler items was to minimize the possibility that subjects would notice the false presuppositions.

37   All subjects returned I wk later and, without reviewing the film clip, answered 20 new questions about the incident. Five of these questions were critical: They were direct questions, shown in Table I, that had been asked a wk earlier in identical form, of only one of the three groups of subjects. The subjects responded to all questions by circling "yes" or "no" on their questionnaires.

Results and Discussion

38   The percentage of subjects responding "yes" to each of the key questions during the final experimental session is shown in Table 1. Overall, [begin page 568] [Table 1 is on page 568 in original article and appears as the last page of this document].

39   [begin page 569] of those who had been exposed to questions including a false presupposition, 29.2% said "yes" to the key nonexistent items; of those who had been exposed to the direct questions, 15.6% said "yes" and of those in the control group, 8.4% said "yes."

40   For each question individually, the type of prior experience significantly influenced the percentage of "yes" responses, with all chi-square values having $p < .05$. Additional chi-square tests were performed to test for the significance of the differences between the pairs of groups. For each of the five questions, the differences were all significant between the control group and the group exposed to false presuppositions, all chi-square values having $p < .025$. Summing over all five questions, a highly significant chi-square resulted, $\chi^2 (5) = 40.79, p < .001$. Similarly, over all five questions, the difference between the group exposed to direct questions and the group exposed to false presuppositions was significant, $\chi^2 (5) = 14.73, p < .025$. The difference between the control group and the group exposed to direct questions failed to reach significance, $\chi^2 (5) = 9.24, p > .05$.

General Discussion

41   We saw that either a strength hypothesis or a construction hypothesis would account for the results of the first experiment in which the presupposition of a true event increased the later assertion that the event had occurred. But only the construction hypothesis explains the comparable results which occur when the presupposition is of false information, as in Experiments 2-4.[1]

42   We need, therefore, to consider the form of a theory of memory for complex visual experiences in which a con-

---

1. It should be emphasized that even though Experiments 2-4 demonstrate support for a construction hypothesis, a strength hypothesis is not necessarily excluded as an explanation for Experiment 1.

structic mechanism plays an integral role. Figure 1 presents a skeleton of this theory that has three major components. The first two components involve acquisition processes, and the third involves retrieval processes.

Acquisition Processes			Retrieval Processes	
Acquisition of Original Experiences		Acquisition of Subsequent Information		
Decision About What to Look At	Integration of Information into some Representation	Integration of new Information into the Representation	Representation of the Altered Representation	Response
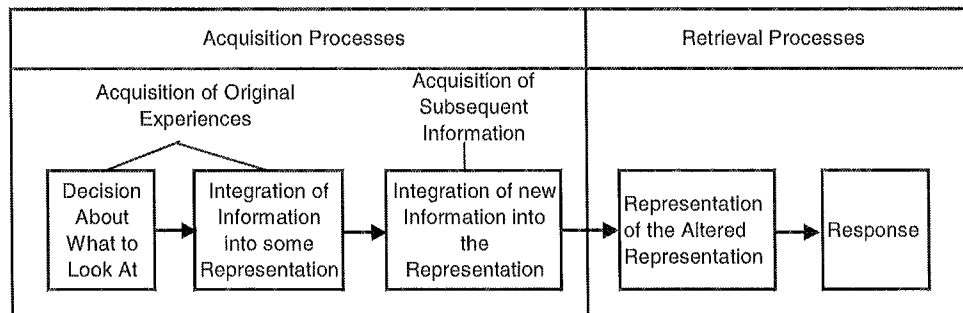
Fig. I. Schematic diagram of the memorial processes. [begin page 570]

Acquisition Processes

43 *Acquisition of the original experience.* When a complex event is experienced, we assume that some of the features of that experience are extracted for arriving at action decisions and/or storage. Early on, the observer must decide to which aspects of the visual stimulus he should attend. Our visual environment typically contains a vast amount of information, and the proportion of information that is actually perceived is very small. The process of deciding to what we attend must consist of a series of decisions, each corresponding to where the next eye fixation should be.

44 *The form of the representation.* Into what form of representation is the newly acquired information integrated? Many views have been suggested. A prominent view is that when a person experiences an event, he organizes and retains knowledge about that event in the form of statements or propositions that can be treated as a labeled graph structure (e.g., Anderson & Bower, 1973; Rumelhart, Lindsay & Norman, 1972). In this view, experience might appear as a collection of points or nodes representing particular concepts or objects, with links between the nodes representing labeled semantic relationships between the particular objects.

45 Other hypotheses about the representation of knowledge are stated in terms of decision routines (e.g., Winograd, 1972); features (e.g., Selfridge & Neisser, 1963); or "mental images" that are isomorphic to the original event (Shepard, 1966). At present, the issue is clearly unresolved. One appealing resolution, however, is that people may use more than one form of representation they may be sufficiently flexible to store information in whichever form is most appropriate to the situation, and they may transform information from one form to another at will. So, for example, human beings may be able to store information in terms of propositions which are then transformed into mental images at the time the information is retrieved.

46 *Acquisition of subsequent information.* However an event may be represented, there is little reason to believe that the representation is accurate; in fact, it may be quite malleable by occurrences other than the event it is supposed to represent. Events or information occurring subsequent (and probably prior) to the original event may alter the representation of that event. One way this might be accomplished is by simply influencing the process of entering new information into the existing memory structure, thereby enhancing, enriching, or otherwise altering that structure. We will refer to the added information as "external" to distinguish it from the information acquired during the initial experience. [begin page 571]

Retrieval Processes

47 Some time after both the initial visual experience and the first interrogation about it, a witness may be quizzed again. For example, after being questioned by the police, a witness may have to testify in court. At this point he must "re-create" from long-term memory, at least that portion of the experience needed to answer a specific

question. Thus, the image may be based both on information acquired during the original experience and external information acquired subsequently. This regenerated image has some internal structure, which may or may not be "visual," but must contain information as to the spatial structure of its referent. Any response which a witness makes is based on this regenerated image.

48   To reiterate, we suggest that information acquired during a complex experience is apparently integrated into some overall memory representation. Subsequent information about that event — for example, that introduced inadvertently via questions containing true or false presuppositions — is also integrated, and can alter the initial representation. When the person is later queried about the original experience, he forms a regenerated image based on the altered memorial representation, and bases his response on that image.

49   In thinking about the present work in relation to some of the existing literature on reconstructive memory, Bartlett's (1932) notions come immediately to mind. Bartlett was one of the first to argue that the way we represent experiences in memory is determined by our permanent knowledge about objects, events, and processes of our experiences. In this view, the new experience is somehow assimilated into the framework of prior experiences. Since Bartlett's work, there has been a lasting interest in the interaction of prior knowledge and present input experiences (cf. Bransford & Johnson, 1972; Dooling & Lachman, 1971). The belief that a person's prior knowledge can wield considerable influence over his recollection of a specific experience is expressed in the recent articles of several noted cognitive psychologists. For example, Rumelhart and Norman (1973) make the point that the "retrieval of an experience from memory is usually a reconstruction which is heavily biased by the person's general knowledge of the world" (p. 450), while Tulving and Thomson (1973) regard "remembering" as "a joint product of information stored in the past and information present in the immediate cognitive environment of the rememberer." (p. 352).

50   The present work extends these notions to include the influence on a to-be-remembered experience of information acquired subsequent to that experience. In the present experiments, the subsequent information was introduced via presuppositions in questions, a technique which is ef- [begin page 572] fective in introducing information without calling attention to it. Obviously, there are many other ways to introduce new information. The experimental manipulation of subsequent information may constitute a useful technique for investigating the interaction of a person's specific experiences and subsequent knowledge related to those experiences.

References

Anderson, J. R., & Bower. G. H. *Human Associative Memory.* Washington, DC: Winston, 1973.

Bartlett, F. C. *Remembering: A study in experimental and social psychology.* London: Cambridge. University Press, 1932.

Bransford, J. D. & Johnson, M. K. Contextual prerequisites for understanding: Some investigations of comprehension and recall. Journal of Verbal Learning and Verbal Behavior, 1972, 11, 7 17-726.

Clark, H. H.. & Haviland. S. E. Psychological processes as linguistic explanation. In D. Cohen (Ed.), *The nature of explanation in linguistics.* Milwaukee: University of Wisconsin Press, in press.

Dooling, D. J., & Lachman, R. Effects of comprehension on retention of prose. *Journal of Experimental Psychology,* 1971, *88,* 216-222.

Harris, R. J. Answering questions containing marked and unmarked adjectives and adverbs. *Journal of Experimental Psychology,* 1973, *97,* 399-401.

Loftus, E. F. Reconstructing memory. The incredible eyewitness. *Psychology Today,* 1974, *8,* 116-119.

Loftus, E. F., & Palmer, J. C. Reconstruction of automobile destruction: An example of the interaction between language and memory. *Journal of Verbal Learning and Verbal Behavior,* 1974, *13,* 585-589.

Loftus, E. F., & Zanni, G. Eyewitness testimony: The influence of the wording of a question. *Bulletin of the Psychonomic*

Society, 1975, *5,* 86-88.

Rumelhart, D. E. Lindsay. P. H., & Norman, D. A. A process model of long-term memory. In E. Tulving & W. Donaldson (Eds.), *Organization of memory.* New York: Academic Press, 1972.

Rumelhart, D. E. . & Norman. D. A. Active semantic networks as a model of human memory. *Proceedings of the Third International Joint Conference on Artificial Intelligence,* Stanford University, 1973.

Shepard, R. N. Learning and recall as organization and search. *Journal of Verbal Learning and Verbal Behavior,* 1966, *5,* 201-204.

Selfridge, 0. G. & Neisser, U. Pattern recognition by machine. In E. A. Feigenbaum & J. Feldman (Eds.), *Computers and thought.* New York: McGraw Hill, 1963.

Tulving, E. & Thomson, D. M. Encoding specificity and retrieval processes in episodic memory. *Psychological Review,* 1973, *80,* 352-373.

Winograd, T. Understanding natural language. *Cognitive Psychology,* 1972, *3,* 119 1.

TABLE 1

PERCENTAGE OF "YES" RESPONSES TO DIRECT QUESTIONS ASKED 1 WK AFTER THE FILM, FOR THE CONTROL GROUP (C), THE DIRECT GROUP (D), AND THE FALSE RESUPPOSITION GROUP (F). ALL QUESTIONS REFERRED TO ITEMS THAT WERE NOT PRESENT

Direct questions	False presupposition questions	Percentage of "yes" responses to direct question 1 wk later[a]			Chi-square	p
		C	D	F		
Did you see a school bus in the film?	Did you see the children getting on the school bus?	6	12	26	8.44	.025
Did you see a truck in the beginning of film?	At the beginning of the film, was the truck parked beside the car?	0	8	22	26.01	.01
Did you see a center line on the country road?	Did another car cross the center line on the country road?	8	14	26	6.26	.05
Did you see a woman pushing the carriage?	Did the woman who was pushing the carriage cross into the road?	26	36	54	8.52	.025
Did you see a barn in the film?	Did you see a station wagon parked in front of the barn?	2	8	18	7.66	.05

[a] Means: C, 8.4; D, 15.6; F, 29.2.

THIRTY YEARS OF INVESTIGATING THE OWN-RACE BIAS IN MEMORY FOR FACES
A Meta-Analytic Review

Christian A. Meissner and John C. Brigham
Florida State University

The current article reviews the own-race bias (ORB) phenomenon in memory for human faces, the finding that own-race faces are better remembered when compared with memory for faces of another, less familiar race. Data were analyzed from 39 research articles, involving 91 independent samples and nearly 5,000 participants. Measures of hit and false alarm rates, and aggregate measures of discrimination accuracy and response criterion were examined, including an analysis of 8 study moderators. Several theoretical relationships were also assessed (i.e., the influence of racial attitudes and interracial contact). Overall, results indicated a "mirror effect" pattern in which own-race faces yielded a higher proportion of hits and a lower proportion of false alarms compared with other-race faces. Consistent with this effect, a significant ORB was also found in aggregate measures of discrimination accuracy and response criterion. The influence of perceptual learning and differentiation processes in the ORB are discussed, in addition to the practical implications of this phenomenon.

> She based her identification on Smith's eyes, which she said were greenish-blue and upon his hands which she said were "light and slender" like the holdup man's. Mrs. McCormick testified that Smith's eyes were "different from most colored people . . . bright and piercing." Smith's defense attorneys then attempted to parry the state's first thrust in the trial. Mrs. McCormick was handed a picture of a man she couldn't identify. It was a picture of David Charles, with shorter hair, taken while he was in Vietnam. Assistant defense attorney Kitchen asked Mrs. McCormick if she had ever made the statement that all Black people look alike. "Yes, I made that statement," Mrs. McCormick said, "and they do to a certain extent, but there's a difference here" (Lickson, 1974, p. 66).

In 1971, five Black men, who became known as the "Quincy Five," were wrongfully indicted for the murder of Khomas Revels during a robbery in Tallahassee, Florida. Although no forensic evidence obtained from the crime scene was ever linked to the men, five White eyewitnesses positively identified them as among the perpetrators. In each of three trials the state argued, "What better evidence can there be than, 'I saw him,' from unprejudiced witnesses? This has been used since time immemorial. This is proof beyond a reasonable doubt. Five eyewitnesses!" (Lickson, 1974, p. 87). Despite the lack of physical evidence against these men, two of the defendants, Dave Roby Keaton and Johnny Frederick, were found guilty on the basis of eyewitness testimony and coerced confessions obtained by investigators. During the third trial involving David

---

Christian A. Meissner and John C. Brigham, Department of Psychology, Florida State University.

Correspondence concerning this article should be addressed to Christian A. Meissner or to John C. Brigham, Department of Psychology, Florida State University, Tallahassee, Florida 32306-1270. Electronic mail may be sent to meissner@psy.fsu.edu.

Charles Smith, hired investigators on the defense team located the three actual perpetrators of the robbery and murder, who became known as the "Jacksonville Three." The Jacksonville men were later brought to trial and convicted based on latent fingerprint evidence and identification of the automobile used in the murder. The Quincy Five were finally exonerated.

At the trial of David Charles Smith, social psychologist Dr. William Haythorn of Florida State University (a colleague of John C. Brigham) was called as an expert witness in rebuttal of the eyewitness misidentifications. Because the only evidence against the Quincy Five was in the form of cross-racial identifications, Haythorn and Brigham set out to locate empirical evidence on the often purported claim that "they [other-race persons] all look alike." However, at the time of this case (c. 1971) only a handful of studies had examined the phenomenon (Berger, 1969; Horowitz & Horowitz, 1938), and only one study had been published in the previous decade (Malpass & Kravitz, 1969). Due in part to this lack of scientific evidence on cross-racial identification, the court prohibited the expert testimony of Haythorn.

Today, three decades later, a plethora of researchers have studied the own-race bias (ORB) in memory for human faces (also referred to as the cross-race effect or other-race effect). Although most now agree that the phenomenon is reliable across cultural and racial groups (Kassin, Ellsworth, & Smith, 1989), there is less consensus about the social and cognitive mechanisms that may govern the effect. Furthermore, little is known regarding variables that might moderate the effect, including those applicable to the eyewitness scenario, such as study time and retention interval. Thus, the goal of the current review and meta-analysis is not only to reconsider the reliability and generalizability of the ORB, but also to evaluate the validity of various theoretical mechanisms previously discussed in the literature and to propose a framework that might best account for the pattern of results across studies. Finally, we discuss the various practical implications of our findings for the legal and criminal justice systems.

Reliability of the ORB Effect

Literature reviews of the ORB have noted the robustness of the phenomenon (Brigham & Malpass, 1985; Chance & Goldstein, 1996), and researchers have endorsed the importance and reliability of the effect in several surveys (Kassin et al., 1989; Yarmey & Jones, 1983). Furthermore, expert witnesses have cited the effect in cases involving disputed cross-race identification (Brigham, Wasserman, & Meissner, 1999; Leippe, 1995), and attorneys have acknowledged the importance of racial interactions in eyewitness identifications (Brigham, 1981; Brigham & WolfsKeil, 1983). Given the source of such endorsements, one might be quick to concede the robust and generalizable nature of the ORB effect. However, it is important to further investigate the particular levels at which reliability might be assessed. For example, (a) Is the effect generally replicable across studies? (b) Is the effect consistent across various racial/ethnic groups? (c) Is the effect significant across different types of memory tasks? and (d) Is the effect reliable across individuals and testing occasions?

OWN-RACE BIAS SYNTHESIS

Replicability Across Studies

The first issue has been explored in several previous meta-analytic reviews of the effect. Bothwell, Brigham, and Malpass (1989) found that roughly 80% of the samples they reviewed demonstrated a significant ORB effect. Overall, effect size estimates from several previous meta-analyses (Anthony, Copper, & Mullen, 1992; Bothwell et al., 1989; Shapiro & Penrod, 1986) have indicated a significant weak-to-moderate effect, accounting for 6% to 11% of the variability across studies. R. C. Lindsay and Wells (1983) also examined the reliability of the effect across 13 studies by way of a vote-counting procedure. Although they asserted that fewer than half of the studies (6 of 13) demonstrated a true ORB effect, their criterion requiring a complete crossover interaction of White and Black participants may have been overly stringent. As Chance and Goldstein (1996) later noted, a large majority of the studies (11 of 13) reviewed by R. C. Lindsay and Wells showed at least some evidence of the effect.

Consistency Across Racial/Ethnic Groups

Several of these reviews have also examined the consistency of the ORB effect across racial/ethnic groups. Whereas Bothwell et al. (1989) found relatively equivalent estimates for both White and Black individuals, Anthony et al. (1992) found that the ORB effect among White participants accounted for 2.5 times the variance than that among Black participants. These inconsistencies could be due to the analysis of slightly different groups of studies. Moreover, both reviews relied on moderately small samples, in meta-analytic terms (number of independent samples: $ks = 28$ and 44; number of participants: $ns = 1,445$ and $1,725$, respectively), increasing the likelihood of significant fluctuations in moderator effects due to the influence of one or more studies.

Generalizability Across Memory Tasks

Most studies documenting the ORB effect have used a standard recognition paradigm in which participants are tested on their ability to discriminate between a subset of faces shown previously (targets) and a subset of novel faces (distractors). Although a handful of studies have utilized some variant of this basic task (Cross, Cross, & Daly, 1971; D. S. Lindsay, Jack, & Christian, 1991; Luce, 1974; Malpass, 1974), some reviewers, such as R. C. Lindsay and Wells (1983), have criticized the literature for not examining performance on other memory tasks, including more applied identification tasks. More recently, however, researchers have responded to this criticism by documenting the effect across a variety of paradigms, including matching tasks (Malpass, Erskine, & Vaughn, 1988) and lineup identification paradigms (Berger, 1969; Brigham, Maass, Snyder, & Spaulding, 1982; Doty, 1998; Fallshore & Schooler, 1995; Platz & Hosch, 1988). In addition, researchers have shown the presence of the effect across other measures of performance such as reaction time (Chance & Goldstein, 1987; Valentine, 1991) and other tasks of forensic relevance including facial reconstruction tasks (Ellis, Davies, & McMurran, 1979) and photo lineup construction by law enforcement officers (Brigham & Ready, 1985).

Reliability Across Individuals and Testing Occasions

A fourth level of reliability, namely consistency of the ORB effect across individual participants, has only recently been examined. In general, memory for human faces has been shown to demonstrate reliable properties when assessed by tools designed to investigate cognitive maturation and/or neurological impairment. For example, Malina, Bowers, Millis, and Uekert (1998) found that the Faces subtest of the Recognition Memory Test (Warrington, 1984) had sufficient internal consistency and reliability (Cronbach's $\alpha = .77$) for clinical use (see also Soukop, Bimbela, & Schiess, 1999). Similarly, the Benton Facial Recognition Test (Benton, Hamsher, Varney, & Spreen, 1983), the Faces subtest of the Wechsler Memory Scale—III (1997), and the Face Recognition subtest of the Kaufman Assessment Battery for Children (Kamphaus, Beres, Kaufman, & Kaufman, 1996) have all produced sizeable reliability estimates ($rs > .75$). Interestingly, and pertinent to the current investigation, several laboratory efforts at demonstrating reliability in a face-recognition task have yielded only moderate reliability estimates (Chance & Goldstein, 1979; Goldstein & Chance, 1980; Malpass et al., 1998; Prospero, Corey, Malpass, Parada, & Schreiber, 1996). Although researchers had taken care to randomly assign faces to recognition sets in controlling for item effects (see Chance & Goldstein, 1979), more deliberate standardization in controlling the memorability of materials and test sets may provide for better estimates of facial memory reliability in future studies.

Although it has largely been assumed that the ORB effect would follow a similar pattern of reliability across testing occasions (namely, moderate-to-large reliability estimates), little research has been available to test this assumption. In a recent study, we (Slone, Brigham, & Meissner, 2000) sought to test the reliability of the ORB effect across an immediate and (2-day) delayed testing occasion. Our results indicated that although participants performed reliably on both own-race and other-race faces, rs (127) $= .56$ and $.44$, $ps < .001$, respectively, the magnitude of the difference between own-race and other-race performance (i.e., the ORB) was only somewhat reliable across the delay, $r(127) = .21$, $p < .05$). Malpass et al. (1998) also recently investigated the reliability of other-race face recognition across two separate testing occasions. Although they found no reliability in performance on other-race faces, $r(11) = .08$, ns, this may have been due, in part, to the small sample of participants ($n = 13$). Their estimate of reliability for same-race recognition was significant, but of moderate size, $r(59) = .36$, $p < .01$. Although left unaddressed by the current meta-analysis, the issue of test–retest reliability in the ORB merits further investigation. Once again, greater care in the standardization of materials across race of face may provide more reasonable estimates of reliability in future studies.

The Search for Social–Cognitive Mechanisms

Thus far, theoretical notions for the ORB have spanned the realms of both social and cognitive mechanisms. Whereas early candidates included the effect of social attitudes and the notion of physiognomic differences between races, more recent hypotheses have involved the potential influence of interracial contact and the notion of a perceptual learning mechanism. Unfortunately, inconsistency has often plagued the literature seeking to verify each theory. Because previous

reviews of the ORB effect have given much attention to the various theoretical positions (Brigham & Malpass, 1985; Chance & Goldstein, 1996; Shepherd, 1981), we provide only a cursory updated description of each approach.

Racial Attitudes

One initial explanation for the ORB effect was that individuals with less prejudiced racial attitudes would be more motivated to differentiate other-race members, when compared with more prejudiced persons. Early research indicated that racial attitudes appeared to influence the degree of stereotypic likeness assigned to other-race members (Secord, Bevan, & Katz, 1956). In addition, early studies examining participants' performance on identification of race/ethnicity (e.g., Jewish vs. non-Jewish) demonstrated that more-prejudiced individuals often performed better than less-prejudiced individuals (Allport & Kramer, 1946; Lindzey & Rogolsky, 1950). However, other studies were not always supportive of the findings (Carter, 1948), and subsequent researchers noted that high-prejudiced performance was likely influenced by a response bias to label more faces as out-group members (Elliott & Wittenberg, 1955).

Within the ORB literature, several early studies demonstrated a small relationship between attitudes toward other-race persons and recognition memory performance (Berger, 1969; Galper, 1973). However, when response bias was taken into account, Dowdle and Settler (cited in Yarmey, 1979) found that racial attitudes were unrelated to memory performance. Similarly, more recent studies have consistently failed to find a relationship between racial attitudes and memory for other-race faces (Brigham & Barkowitz, 1978; Lavrakas, Buri, & Mayzner, 1976; Platz & Hosch, 1988; Slone et al., 2000; Swope, 1994). However, racial attitudes are related to another factor thought relevant to recognition of other-race faces, namely, amount of interracial contact. A number of studies have found that those with more prejudiced attitudes report less contact with other-race members (Brigham, 1993; Brigham & Barkowitz, 1978; Brigham & Meissner, 2000; Brigham & Ready, 1985; Slone et al., 2000; Swope, 1994).

Physiognomic Homogeneity

A second possibility for the ORB effect involves possible group differences in the inherent memorability of faces, such that faces of some races might show less physiognomic variability among group members when compared with other races. However, researchers examining this hypothesis have generally found little support for its validity. For example, Goldstein (1979) found no differences in physiognomic variability among Japanese, Black, and White faces. Additionally, several studies have demonstrated that latency and accuracy of same–different judgments do not differ across race of participant or race of face (Goldstein & Chance, 1976, 1978). Finally, within-race rated similarity has shown, at best, only an inconsistent relationship to perception by own-race and other-race individuals, leading Goldstein and Chance (1979) to conclude that, overall, there is little "compelling evidence for the homogeneity hypothesis" (p. 111). We should note that although physiognomic homogeneity may not be responsible for the ORB memory effect, a number of studies have indicated that different physiognomic facial features may be more appropriate for discriminating between faces of

certain races (Ellis, Deregowski, & Shepherd, 1975; Shepherd, 1981; Shepherd & Deregowski, 1981).

Interracial Contact

A number of researchers have posited that the quality or quantity of interracial contact may play a vital role in the degree of ORB demonstrated by any particular individual. For example, researchers have proposed that increased contact with other-race individuals may increase memory performance by (a) reducing the likelihood of stereotypic responses and increasing the likelihood that individuals may look for more individuating information (Malpass, 1981; Shepherd, 1981), (b) influencing individuals' motivation to accurately recognize other-race persons through associated social rewards and punishments (Malpass, 1990), or (c) reducing the perceived complexity of unfamiliar other-race faces (Goldstein & Chance, 1971). Two major approaches to investigating contact are to examine groups of individuals differing in their degree of other-race contact or to assess individuals' self-reported contact with other-race persons.

With regard to the former approach, several early studies demonstrated that adolescents and children living in integrated neighborhoods better recognized novel other-race faces than did those living in segregated neighborhoods (Cross et al., 1971; Feinman & Entwisle, 1976). Other more recent studies have also shown evidence of the influence of contact in samples of White and Black individuals from Great Britain and Africa (Carroo, 1986; Chiroro & Valentine, 1995; Wright, Boyd, & Tredoux, 1999). Finally, a novel application of the contact hypothesis was recently conducted by Li, Dunning, and Malpass (1998) who demonstrated that White "basketball fans" were superior to White "basketball novices" in recognizing Black faces. Given that the majority of professional basketball players are Black, this effect was predicted on the basis of the fans' experience in differentiating individual players. It is interesting to note that not all studies have found the predicted relationship between high-contact and low-contact groups. Burgess (1997) found only a small effect of contact on the performance of Southern (Florida) and Northern (Maine) American samples of White individuals. Similarly, Ng and Lindsay (1994) found little support for the influence of contact on the performance of Canadian and Singapore samples.

In a number of other studies, researchers have assessed the relationship between memory for other-race faces and individuals' self-reported experience with other-race persons. Whereas early studies generally failed to find a significant relationship (Berger, 1969; Brigham & Barkowitz, 1978; Cross et al., 1971; Malpass & Kravitz, 1969), numerous studies over the past several decades have found at least some evidence of the relationship in both recognition tasks (Byatt & Rhodes, 1998; Carroo, 1986, 1987; Lavrakas et al., 1976; Li et al., 1998; D. S. Lindsay et al., 1991; Slone et al., 2000; Swope, 1994; Wright et al., 1999) and more applied lineup identification paradigms (Brigham et al., 1982; Platz & Hosch, 1988). This curious pattern of results over time will be further examined in the current meta-analysis. It is possible that the precision and validity of measures used to assess interracial contact have improved over the years. Alternatively, as Chance and Goldstein (1996) posited, a cohort effect may exist such that opportunities for interracial contact have increased following the desegrega-

tion and civil rights movements of the 1960s and 1970s, allowing for a greater range in the degree of interracial contact in recent years.

Perceptual Learning

As reviewed in the previous section, a fair degree of empirical support exists for the notion that interracial contact has some influence on the magnitude of the ORB. However, researchers are still attempting to elucidate the specific cognitive mechanisms through which contact might actuate this influence, and to model their effects in more formal ways. The most popular general approach is likely that of *perceptual learning*. As historically defined by Gibson (1969), perceptual learning involves "an increase in the ability to extract information from the environment, as a result of practice and experience with stimulation coming from it" (p. 3). Numerous reviews have been written concerning the various mechanisms likely to underlie the phenomenon (Ahissar & Hochstein, 1998; Proctor & Dutta, 1995; Walk, 1978), and most note the important role of Gibson's notion of *differentiation*, defined as focused attention directed toward invariant cues that provide the best bases for discriminations within a given stimulus set. More recent work by Haider and Frensch (1996, 1999) has furthered Gibson's notion by demonstrating that perceptual skill involves learning to distinguish between "task-relevant" and "task-redundant" information. Thus, increases in accuracy and speed of processing appear to reflect the extent to which individuals have knowledge of, and provide attention to, the appropriate (invariant) features of the stimulus.

Such an encoding-based effect has been documented in a variety of perceptual skill domains, including chess (Reingold, Charness, Pomplun, & Stampe, in press), bird watching (K. E. Johnson & Mervis, 1997, 1998), sports (Helsen & Pauwels, 1993; Shea & Paull, 1996), radiology (Christensen et al., 1981; Lesgold et al., 1988; Myles-Worsley, Johnston, & Simons, 1988), and even chicken sexing (Biederman & Shiffrar, 1987). It is possible that perceptual learning might also be responsible for the ORB phenomenon. For example, individuals may be able to discriminate own-race faces more accurately due to their use of appropriate (invariant) aspects of the face. On the other hand, cues used for own-race faces may not be appropriate when attempting to remember other-race faces, and thus performance would worsen when attempting to discriminate such unfamiliar stimuli. A handful of studies have investigated this notion of perceptual learning from a discrimination training perspective. Other research within this general framework has attempted to identify various aspects of the face that might be deemed "task-relevant" when recognizing own-race versus other-race faces and to provide evidence in support of more formal models of the ORB.

Discrimination training. Some researchers in the face memory domain have directly investigated the perceptual learning hypothesis by providing individuals with discrimination training on own-race and other-race faces. Although training seems to have no effect on improving own-race recognition (Malpass, 1981), there is some evidence that training may reduce the ORB, at least in the short run. For example, Malpass, Lavigueur, and Weldon (1973) attempted to improve recognition memory for own-race and other-race faces by either verbal or visual training tasks. Although verbal training showed no effect on recognition, a

relatively short visual training task (1 hr) produced a significant reduction in the magnitude of the ORB. Lavrakas et al. (1976) also investigated the effects of training by presenting participants with a concept learning task. Post-training recognition performance demonstrated significant improvement on other-race faces for individuals in the concept learning conditions compared with the unchanged performance of individuals in a control condition. However, when participants in all conditions were tested again 1 week later, the performance of trained and untrained participants on other-race faces was no different. Finally, E. S. Elliott, Wills, and Goldstein (1973) investigated the influence of paired associate discrimination training in reducing the magnitude of the ORB. Whereas participants in the no-training and own-race training conditions displayed the typical ORB effect, those in the other-race training condition demonstrated significant improvement in recognition accuracy for other-race faces.

Configural–featural hypothesis. Although relatively short-lived effects of discrimination training have been found, other researchers have sought to identify the various cognitive processes that might differentiate own-race and other-race face recognition. One notable advance in the face memory literature has involved work on the *face inversion effect,* the finding that inverted (upside-down) photos of faces are identified more poorly than inverted photos of other objects. In early work on this effect, Yin (1969) concluded that face recognition was the product of a unique system, different from systems responsible for recognizing other kinds of visual stimuli. In contrast to this "neural specialization" hypothesis, Diamond and Carey (1986) proposed that perceptual learning might be operating in face recognition. In several experiments they showed that the inversion effect was not unique to faces, but rather occurred when participants had a great deal of experience with the stimulus materials. Inversion appeared to disrupt the effectiveness with which individuals were able to encode stimuli that were highly familiar to them. This, they claimed, stemmed from experienced participants' reliance on configural (or relational) properties of the stimulus. Novice participants, on the other hand, relied on only the featural (or isolated) aspects of the face that were less influenced by inversion. A number of subsequent studies have supported this general configural–featural hypothesis (see Farah, Wilson, Drain, & Tanaka, 1998).

The notion of expertise and configural processing has also been applied to the ORB effect. In particular, Rhodes, Brake, Taylor, and Tan (1989) proposed that greater experience with own-race faces would lead to a larger inversion effect, due to an increased reliance on configural information. The encoding of other-race faces, on the other hand, should not be as influenced by inversion due to the featural aspects that are relied on. As hypothesized, Rhodes et al. observed that own-race faces were significantly more susceptible to inversion than other-race faces for measures of both reaction time and accuracy. However, several other studies have observed either no interaction of inversion with the ORB (Buckhout & Regan, 1988) or larger inversion effects on other-race faces (Valentine & Bruce, 1986). Given the various methodological differences across studies, further empirical and theoretical work on the significance of inversion effects in the ORB would be valuable.

Finally, Fallshore and Schooler (1995) examined whether such perceptual expertise might also be involved in the *verbal overshadowing effect,* the finding

that generating a verbal description of a face significantly impairs subsequent identification accuracy (see Meissner & Brigham, in press, for a meta-analytic review). Specifically, they hypothesized that requesting participants to provide a description of a same-race face might cause significant declines in recognition performance by (a) forcing participants to rely on the featural (more verbalizable) aspects of the face and (b) disrupting the configural (less verbalizable) memory trace that was originally encoded. Performance in cross-race identification, however, was predicted not to show the overshadowing effect due to individuals' reliance on featural aspects when encoding other-race faces. Consistent with their hypotheses, Fallshore and Schooler found that although participants' recognition performance on same-race faces demonstrated the overshadowing effect (a 47% decrement in performance when verbal descriptions were given), other-race faces showed no such decline in performance.

"Face space" models. Although the configural–featural hypothesis has received much attention, other researchers have examined the particular manner in which faces might be represented in memory. Likely the most ambitious work involves that of Valentine and his colleagues (Valentine, 1991; Valentine & Bruce, 1986; Valentine & Endo, 1992) in the development of an exemplar-based model of facial memory. Although Valentine and colleagues conceded the notion of a configural–featural distinction in the type of facial features that individuals may encode, they disputed Diamond and Carey's (1986) proposal that a fundamental change in the underlying processing strategy occurs under inversion (Valentine, 1988; Valentine & Bruce, 1988). Rather, Valentine (1988) proposed that, in conjunction with the notion of *schema theory* pioneered by Goldstein and Chance (1980), an exemplar-model reflecting "the acquisition of knowledge of how faces vary" may account for the effects of inversion, race, and distinctiveness (Valentine, 1988, p. 485).

Generally speaking, Valentine's (1991) multi-dimensional space (MDS) framework holds that the representational system may be thought of as a hypothetical space in which faces are stored based on various dimensions representing features or sets of features. The model posits that these dimensions are based on an individual's prior experience with the stimulus set and thus are best suited for representation of own-race faces, due to a reliance on appropriate featural and/or configural information. As a result of this encoding, own-race faces are spread more evenly throughout the MDS and are better individuated from one another at retrieval. Conversely, other-race faces are poorly represented (and, thus, more tightly clustered in the MDS) due to the encoding of less appropriate featural and/or configural information. Valentine's (1991) model also posits, however, that with increasing experience, other-race faces may be better represented once the relevant (invariant) aspects of other-race faces are learned.

In a test of the MDS framework, Chiroro and Valentine (1995) examined the effects of race, typicality, and level of perceptual experience within the cross-race paradigm. Although the influence of rated distinctiveness on recognition of own-race faces had been widely known (Brigham, 1990; Hosie & Milne, 1995), the manner in which it might interact with race and perceptual experience had not been investigated. Based on the assumptions of the MDS model, Chiroro and Valentine predicted that only individuals who had considerable previous experience with other-race faces (high-contact) would demonstrate distinctiveness ef-

fects for both own-race and other-race faces. This was due largely to the notion that such individuals should be able to distinguish between typical and distinctive other-race faces based on features they had extracted through prior experience. In contrast, low-contact individuals were predicted to demonstrate no differences in performance on the distinctiveness dimensions of other-race faces. Overall, their results indicated the predicted four-way interaction such that distinctiveness effects for low-contact individuals were confined to own-race faces. On the other hand, high-contact individuals demonstrated significant effects of distinctiveness regardless of the race of the face.

Race-feature hypothesis. An alternative to Valentine's (1991) MDS model was proposed by Levin (1996) in explaining the paradoxical effect that individuals are slower at classifying the race of an own-race face compared with that of an other-race face. This other-race classification advantage (ORCA) was observed by Valentine and Endo (1992) and was explained as resulting from strong activation due to the high-density cluster of other-race faces in the representational system (MDS). Levin (1996) proposed an alternative to this explanation in which the ORCA was said to arise from a "facilitated classification process" (p. 1366). In particular, Levin suggested that other-race faces were more quickly classified due to an automated process in which race-specific coding is performed without regard for other individuating information, which is largely ignored.

In testing this race-feature hypothesis, Levin (1996) observed that participants demonstrating a large ORB in recognition memory also demonstrated a large ORCA when compared with other individuals (see also Levin & Lacruz, 1999). Given that the ORB observed was driven largely by false alarm responses to other-race faces, Levin argued that participants' coding of race alone was insufficient to discriminate between other-race faces, leading to a tendency to respond "seen before" during test. Levin further proposed that individuals having greater experience with other-race persons would be less likely to generate the race-feature response, but instead would initially seek out individuating information for later use. Although he did not test this possibility, Levin's observation is analogous to that of skill differences in the "basic level" categorization effect (K. E. Johnson & Mervis, 1997, 1998; Tanaka & Taylor, 1991). Namely, whereas novices respond to stimuli most quickly based on a basic level categorization (e.g., bird), experts respond just as quickly at the basic, subordinate (e.g., wren), and even sub-subordinate levels (e.g., Carolina wren). Thus, experts' conceptual knowledge of domain-relevant features appears to allow them faster access to multiple levels of identification. Similarly, individuals with more experience with other-race faces may have faster access to identity information by way of their conceptual knowledge of individuating features.

Taken together, a perceptual learning approach to understanding the ORB has considerable potential for explaining its cognitive origins. The focus on encoding-based processes within the configural–featural and race-feature hypotheses may stimulate future empirical and theoretical progress. In addition, the representational model put forth by Valentine and colleagues (Valentine, 1991; Valentine & Endo, 1992) has provided a testable framework within which both general and effect-specific approaches to memory for faces may interact. The current meta-analysis was designed to aid researchers in further exploring perceptual learning

aspects of the ORB by providing aggregate estimates of the effect across several performance measures.

Meta-Analysis

The present review of the ORB paradigm has yielded many testable hypotheses concerning both the general reliability of the effect and the various mechanisms posited for its occurrence. Our meta-analysis took the approach advocated by Hedges and Olkin (1985) in which a mean weighted effect size for the sample of studies was initially calculated, followed by prediction of effect size based on moderating variables (see B. T. Johnson, Mullen, & Salas, 1995, for a discussion of various approaches). In particular, we were interested in examining ORB effect size estimates for basic measures of hits (correctly identifying a face as "old") and false alarms (incorrectly identifying a face as "old"), as well as aggregate signal detection estimates of discrimination accuracy (the standardized distance between the means of the "new" and "old" distributions) and response criterion (the level of familiarity necessary for an individual to categorize a given stimulus as "old" vs. "new"; for a review of signal detection theory, see Green & Swets, 1966). Second, in testing the validity of several theoretical mechanisms posited in the literature, we also provide estimates of the influence of racial attitudes and self-rated interracial contact on other-race memory performance, as well as an estimate of the correlation between attitudes and contact as measured across studies. Finally, in addition to overall effect size analyses, eight moderating variables (described below) are examined across the four performance measures.

Method

Studies

A total of 91 independent effect sizes described in 39 research articles were located, representing the responses of 4,996 participants. Of the 39 research articles, 6 (15%) were unpublished manuscripts or theses/dissertations. Studies were obtained using several methods, including (a) searches of *PsycINFO, Sociofile,* and *Dissertation Abstracts* databases and using the key words "face memory," "face recognition," and "face identification" along with the key words "race" and "ethnicity"; (b) cross-referencing with the three previous meta-analyses (Anthony et al., 1992; Bothwell et al., 1989; Shapiro & Penrod, 1986) and various reviews on the effect (Brigham & Malpass, 1985; Chance & Goldstein, 1996; R. C. Lindsay & Wells, 1983); and (c) contact with colleagues in the field who may have had knowledge of fugitive literature that had neither been published nor presented at a conference.

Inclusion–Exclusion Criteria

To be included in the analysis, studies must have involved a within-subjects test of participants' memory for own-race and other-race faces. The statistical difference in performance on these two sets of stimuli for each participant is defined as the ORB. Note that, in contrast to several previous meta-analyses (Anthony et al., 1992; Bothwell et al., 1989), studies that involved only a single race of participants were included in addition to studies that involved races other than Whites and Blacks. Reasons for excluding studies involved (a) the lack of sufficient data from which to compute an effect size (Bruce, Beard, & Tedford, 1997; Caroo, 1988; Horowitz & Horowitz, 1938; Luce, 1974; Malpass,

1988), (b) the use of a between-subjects design and analysis (Caroo, 1986; E. S. Elliot, Wills, & Goldstein, 1973), or (c) the implementation of various methodological procedures that might obscure interpretation of the effect size estimate, such as unequal presentation rates for own-race and other-race faces (Byatt & Rhodes, 1998; Doty, 1998; Goldstein & Chance, 1985; Lavrakas et al., 1976; Padgett, 1997; Valentine & Bruce, 1986).

Coded Variables

Based on the suggestions of Lipsey (1994), moderator variables were selected by way of three general categories of study descriptors.[1] First, we examined variables that were of substantive experimental and applied interest in characterizing the reliability and generality of the ORB effect, including the race of the participant and the type of memory task used. Fifty-six percent of the samples were reported as White, and 32% were reported as Black. The remaining 12% of samples included individuals of Arab/Turkish, Asian, and Hispanic origin. The majority (91%) of studies used a recognition paradigm, whereas 9% of studies used a (simultaneous and target-present) lineup identification task. Briefly, recognition paradigms involve presenting participants with a set of faces that they must later recognize from a group of "old" and "new" faces. Identification paradigms are generally more applicable to the eyewitness situation and involve presenting participants with a single face (either from a photograph or a short video) that they must later identify from a group (or photo lineup) of 6–8 similar faces.

Second, we assessed methodological or procedural aspects of studies such that we might identify possible sources of distortion involving boundary conditions under which the ORB might be observed. Such variables included (a) whether test stimuli were identical (72%) or different (28%) from those used at study, (b) whether races of face were presented and tested in a blocked (19%) or mixed (81%) fashion, (c) the amount of time participants were permitted to study individual faces (minimum = 0.12 s; maximum = 4 min; median = 3 s), and (d) the length of the retention interval between study and test phases (minimum = immediate; maximum = 3 weeks; median = 2 min).

Finally, we also considered other extrinsic study characteristics, including the date of publication or presentation and whether the effect size estimate was taken from a published or unpublished manuscript. Of the studies included for analysis, 27% were published in the 1970s, 33% in the 1980s, and 40% in the 1990s. Fifteen percent of these studies were unpublished and took the form of a conference presentation or a thesis/dissertation.

Measure of Effect Size

Our measure of effect size for the performance variables (i.e., hits, false alarms, and discrimination accuracy) was a single sample estimate equivalent to Hedge's g^U. This effect size was computed simply as the mean difference between own-race and other-race performance divided by the sample standard deviation, or

$$g = (\mu_{own} - \mu_{other})/S_D \qquad (1)$$

To control for skewness in estimating the true population parameter, g was transformed to g^U by way of Equation 2:

---

[1]To assess the reliability of coding study moderator variables, two raters generated independent codings for each variable across studies. Rate of agreement across all variables ranged between 93% and 100%.

OWN-RACE BIAS SYNTHESIS

$$g^{U} = c(m)*g, \tag{2}$$

where

$$c(m) = 1 - (3/[(4*df) - 1]). \tag{3}$$

To assess the influence of both attitudes and contact on the ORB, as well as the correlation between the two measures across studies, r coefficients were recorded for each independent sample, after which r was transformed to Fisher's Z_r by way of Equation 4:

$$Z_r = .5*\log_e[(1 + r)/(1 - r)]. \tag{4}$$

All formulae were obtained from Rosenthal (1994). Effect sizes demonstrating the ORB will be positive for measures of hits, discrimination accuracy, and response criterion, and negative for the measure of false alarms. Likewise, positive estimates for the racial attitude and interracial contact measures indicate that positive attitude toward and increased contact with other-race individuals leads to better performance on other-race faces.

Results

Weighted Effect Size Analyses

To examine the pattern of effect sizes for each measure, estimates were weighted as a function of their independent sample sizes, after which the results were analyzed across studies. For each measure, the mean weighted effect size (g^{U}) is presented, in addition to a test of the significance of the estimate (Z), and the associated 95% confidence intervals.

Hits and false alarms. The mean weighted effect size for the proportion of hit responses across studies ($k = 74$) demonstrated a significant ORB, $g^{U} = .24$, $Z = 15.43$, $p < .001$, with 95% confidence intervals of .21 and .27. In practical terms, an odds-ratio analysis indicated that participants were 1.4 times more likely to correctly identify a previously viewed own-race face when compared with performance on other-race faces. For false alarm responses, the mean weighted effect size across studies ($k = 53$) also indicated a significant ORB, $g^{U} = -.39$, $Z = 22.24$, $p < .001$, with 95% confidence intervals of $-.42$ and $-.35$. Participants were 1.56 times more likely to falsely identify a novel other-race face when compared with performance on own-race faces.

Taken together, these results illustrate a "mirror effect" pattern in which other-race faces receive a lower proportion of hits and a higher proportion of false alarms when compared with own-race faces (Figure 1). The mirror effect has been termed a "regularity" of recognition memory and has been demonstrated for such variables as frequency, distinctiveness, and study time (see Glanzer & Adams, 1985, 1990). Although the theoretical mechanisms of this effect are often debated between models (Glanzer & Adams, 1990; Hintzman, 1988; Hirshman, 1995; McClelland & Chappell, 1998; Shiffrin & Steyvers, 1997), many studies have shown that the aggregate measure of discrimination accuracy is generally influenced when mirror effects are observed. Other researchers have noted changes in response criterion estimates as well; however, substantial differences in discrimination accuracy between stimuli must be present for the criterion effect to be observed (Hirshman, 1995; McClelland & Chappell, 1998). Hence, we were
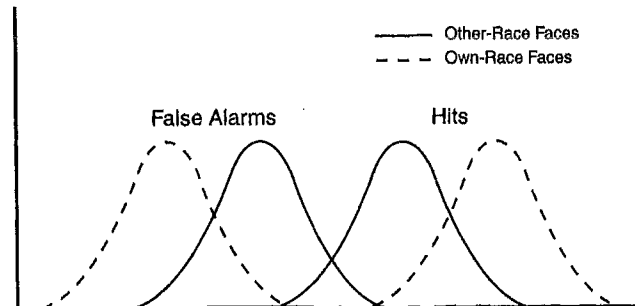
Figure 1. "Mirror-effect" pattern demonstrated in hit and false alarm responses to own-race and other-race faces.

interested to see whether ORB differences would occur only on estimates of discrimination accuracy, or on estimates of response criterion as well.

Discrimination accuracy. The mean weighted effect size for the measures of discrimination accuracy across studies ($k = 56$) was $g^U = .82$, a significant ORB, $Z = 42.32, p < .001$, with 95% confidence intervals of .78 and .85. Overall, the ORB in discrimination accuracy accounted for 15% of the variability across studies, and participants were 2.23 times more likely to accurately discriminate an own-race face as new versus old when compared with performance on other-race faces.

Response criterion. Unfortunately, only six studies ($k = 14$) actually calculated a response criterion measure across participants. Of the 14 independent samples, 11 demonstrated a significant ORB effect ($\alpha = .05$) such that other-race faces yielded a more liberal criterion when compared with performance on own-race faces. The remaining 3 samples demonstrated nonsignificant patterns. To further assess this effect, a studywise response criterion analysis was conducted in which the mean hit and false alarm rates for each study were used to calculate a response criterion estimate (see Macmillan & Creelman, 1990). The mean weighted effect size for the estimates of response criterion across studies ($k = 49$) was $g^U = .30$, a significant ORB, $Z = 17.91, p < .001$, with 95% confidence intervals of .26 and .33. Overall, this small effect of response criterion in the ORB accounted for only 1% of the variability across studies and indicated that own-race faces generally yielded a more conservative criterion when compared with performance on other-race faces.

In summary, the pattern of results for discrimination accuracy measures was consistent with the mirror effect pattern that was observed in the hit and false alarm responses. Given the significant size of the discrimination accuracy effect, the presence of a response criterion effect in the ORB was expected (Hirshman, 1995). A recent model of recognition memory proposed by McClelland and Chappell (1998) provided an account of this pattern of results by simulating the process of differentiation (Gibson, 1969). As discussed previously, differentiation has been implicated in the various perceptual learning approaches to explaining the ORB. In the Discussion section, we consider the merits of McClelland and Chappell's model and its theoretical implications for the ORB.

Racial attitudes. Researchers have long posited that attitudes toward other-race persons may be responsible for the ORB in face memory. However, as noted, empirical results have not generally supported this notion. To assess the validity of this hypothesis, we examined the pattern of correlations between racial attitudes and performance on other-race faces across studies ($k = 14$). The mean weighted effect size across studies indicated no significant relationship, $Z_r = -.01$, $Z = .25$, with 95% confidence intervals of $-.08$ and .06. Hence, there appears to be no evidence of a direct influence of racial attitudes on the ORB.

Interracial contact. Researchers have also posited that interracial contact should influence the degree of ORB demonstrated by any given individual. To assess this relationship across studies, we examined the pattern of correlations between self-rated interracial contact and discrimination of other-race faces ($k = 29$). The mean weighted effect size across studies demonstrated a significant relationship, $Z_r = .13$, $Z = 5.34$, $p < .001$, with 95% confidence intervals of .08 and .18. Overall, contact appears to play a small, yet reliable, mediating role in the ORB, accounting for approximately 2% of the variability across participants. This seemingly weak relationship between self-rated contact and the ORB may be due to limitations in the range of variability present in such measures. Future studies may wish to further explore alternative methods of assessing interracial contact.

Attitude–contact relationship. As noted previously, we have found evidence of a relationship between attitudes toward other-race persons and self-rated contact in our lab. It is conceivable that although individuals' attitudes have no direct influence on their memory for other-race faces, racial attitudes may yet play a mediating role by way of their relation to individuals' social experience with other-race persons. The mean weighted effect size between interracial attitudes and contact across studies ($k = 10$) demonstrated a significant relationship, $Z_r = .36$, $Z = 11.42$, $p < .001$, with 95% confidence intervals of .30 and .42. In general, individuals with more positive attitudes toward other-race persons tend to rate themselves as experiencing more interracial contact when compared with individuals with more negative attitudes.

Moderator Effects

A test of the homogeneity of variances across the sample of weighted effect sizes (hit, false alarm, discrimination accuracy, and response criterion measures) indicated a significant degree of variability, exceeding that expected on the basis of sampling error alone, $Qs > 1,000$, $ps < .001$. Thus, the design moderators discussed earlier were used to predict the variability across the sample of effect sizes. A weighted least-squares regression analysis (Hedges, 1994) was conducted for each measure across the three sets of moderator variables (i.e., reliability and generalizability, methodological characteristics, and extrinsic study factors). Effect sizes in the analysis were weighted as a function of their sample size. Due to the sensitivity of this fixed-effects analysis, we took a more conservative approach

MEISSNER AND BRIGHAM

and discuss only those moderator effects with $Z_j \geq 3.30$ or $\alpha = .001$.[2] Significant effects resulting from this criterion yielded semipartial correlations (r_s) ranging in magnitude from .11 to .33. Table 1 provides a summary of moderator effects (Z_j) across the four performance measures.

Reliability and generalizability. The first set of moderators assessed whether the ORB was reliable across racial/ethnic groups and whether the effect was generalizable to the type of memory task. Similar to that of Anthony et al. (1992), results indicated that White participants demonstrated a significantly larger ORB when compared with Black participants with regard to the measure of discrimination accuracy, $Z_j = 6.91, p < .001$. This effect appeared to stem largely from differences in the magnitude of false alarm responses, $Z_j = 9.50, p < .001$. However, Whites and Blacks did not differ in the magnitude of the ORB on either proportion of hits or estimates of response criterion, Z_js $\leq .79$. White participants also demonstrated a significantly larger ORB when compared with participants grouped in the "other" racial/ethnic category. This effect was observed reliably in hit, false alarm, and response criterion estimates, Z_js $\geq 8.14, ps < .001$. However, the analysis of discrimination accuracy was not significant, $Z_j = 1.13$. Mean weighted effect sizes for each racial/ethnic group across the four performance measures are displayed in Table 2.

Analysis of the effect sizes found in recognition versus lineup identification paradigms yielded no significant difference with regard to the measure of false alarm responses, $Z_j = 1.55$. However, there was a tendency for studies using an identification paradigm ($g^U = .45$) to yield a larger ORB for proportion of hits when compared with studies using a recognition paradigm ($g^U = .22$), $Z_j = 2.76$, $p < .01$. Nevertheless, it is evident that the ORB effect is generalizable to both recognition and lineup identification tasks. As only a small proportion (9%) of the samples involved the use of an identification task, future studies utilizing the lineup paradigm would be valuable.

Methodological characteristics. The second set of moderators examined various methodological aspects that might influence the magnitude of effects observed across studies. First, studies were coded for whether they utilized the identical or different facial photographs at study and test and for whether the presentation of stimuli was mixed or blocked by race/ethnicity. Results indicated that the type of stimulus (i.e., identical vs. different) significantly influenced estimates of the ORB on the proportion of hits and estimates of response criterion, Z_js $\geq 3.42, ps < .001$. This effect of stimulus type was also apparent in the proportion of false alarms, $Z_j = 3.27, p < .01$, though not at the $\alpha = .001$ level.

---

[2] This conservative criterion ($\alpha = .001$) for study moderators was chosen due to the sensitivity of the "fixed effects" analysis. Given the exploratory nature of our investigation, we felt that such a criterion might allow us to examine a range of variables that would likely be replicable under direct empirical investigation. A more conservative, "random effects" model was also run on the sample of studies (see Raudenbush, 1994). Results indicated that White participants yielded a significantly larger own-race bias (ORB) on false alarm responses when compared with both Black, $Z = 2.50$, $p < .05$, and other racial/ethnic participants, $Z = 2.45, p < .05$. White and other participants also exhibited a significant difference in the response criterion estimates, $Z = 2.13, p < .05$. Additionally, limiting the amount of study time significantly increased estimates of the ORB on aggregate measures of discrimination accuracy, $Z = -2.19, p < .05$. No other moderator effects were found to be significant, Zs ≤ 1.36.

OWN-RACE BIAS SYNTHESIS

Table 1

Influence of Moderator Variables (Z_j) Across Measures of Hits, False Alarms, Discrimination Accuracy, and Response Criterion

Moderator	Hits ($k = 74$)	False alarms ($k = 53$)	Discrimination accuracy ($k = 56$)	Response criterion ($k = 49$)
Race/ethnicity of participant				
White vs. Black	0.79	9.50***	6.91***	0.12
White vs. Other	8.14***	10.36***	1.13	11.55***
Task				
Identification vs. recognition	2.76	1.55	—	—
Stimulus				
Photo vs. face	3.42***	3.27	0.16	7.34***
Order of study				
Blocked vs. mixed	4.47***	1.13	5.49***	0.81
Study time (in seconds)	0.60	2.65	11.70***	1.19
Retention interval (in minutes)	0.45	0.86	1.55	7.17***
Year of study	2.50	5.46***	10.50***	0.50
Status of study				
Published vs. unpublished	0.69	0.63	1.74	1.59

***$p < .001$.

In addition, presentation of stimuli (i.e., blocked vs. mixed) significantly influenced ORB estimates of discrimination accuracy, $Z_j = 4.47, p < .001$, largely as a function of the proportion of hits, $Z_j = 5.49, p < .001$. As displayed in Table 3, the pattern of weighted means demonstrated that significantly larger ORB effects were observed if facial photographs were altered from study to test and if the presentation of faces was blocked by race/ethnicity.

    Studies were also coded for the length of time participants studied each target face (in seconds), and the length of the retention interval between study and test phases of the experiment (in minutes). Results indicated that the amount of study time influenced estimates of the ORB on measures of discrimination accuracy, $Z_j = 11.70, p < .001, r_s = -.29$. The direction of the effect indicated that reducing the amount of study time for each face significantly increased the magnitude of the ORB, largely as a result of an increase in the proportion of false alarm responses to other-race faces, $Z_j = 2.65, r_s = -.09$. This effect of exposure time is similar to the findings of Anthony et al. (1992) across their meta-analytic

Table 2

Weighted Effect Size Estimates (g^U) on Performance Measures as a Function of Race/Ethnicity of Participant

Race/ethnicity of participant	Hits	False alarms	Discrimination accuracy	Response criterion
Whites	0.35	−0.62	1.06	0.38
Blacks	0.32	−0.15	0.66	0.32
Others	0.04	−0.22	0.74	−0.21

MEISSNER AND BRIGHAM

Table 3
Weighted Effect Size Estimates (g^U) on Performance Measures as a Function of Changes in Stimuli and Order of Study by Race of Face

Moderator	Hits	False alarms	Discrimination accuracy	Response criterion
Stimuli at study and test				
Identical	0.20	−0.36	0.76	0.14
Different	0.37	−0.48	0.82	0.53
Order of study by race of face				
Mixed	0.21	−0.38	0.79	0.30
Blocked	0.45	−0.45	1.18	0.37

sample of White participants. Length of the retention interval had a significant influence on the size of the ORB across estimates of response criterion, $Z_j = 7.17$, $p < .001$, $r_s = .18$. The direction of the effect indicated that lengthening the retention interval induced more liberal responding to other-race faces.

Extrinsic study factors. As a final set of moderator variables, effect sizes were coded for whether they had been taken from a published or unpublished manuscript and for the date of the manuscript's publication or presentation. Results indicated no significant differences in the magnitude of effect sizes taken from published and unpublished manuscripts, Z_js ≤ 1.74. However, a rather interesting effect was found for date of study, most significantly across measures of false alarm and discrimination, Z_js ≥ 5.46, $ps < .001$. Weighted means for each decade are presented in Table 4. It appears that whereas the size of the ORB has significantly decreased over time for measures of discrimination accuracy, $r_s = −.22$, and proportion of hits, $r_s = −.06$, it has significantly increased over time for the proportion of false alarms, $r_s = −.17$. Curiously, this effect does not hold for estimates of response criterion.

Influence of date of study on estimates of attitude and contact. With regard to estimates of racial attitude and interracial contact, we also assessed the effect of date of study on the magnitude of effects observed. Whereas the estimated influence of racial attitudes on recognition of other-race faces has significantly decreased over the past 3 decades, $Z_j = 16.67$, $p < .001$, $r_s = −.46$, the influence of interracial contact on recognition has significantly increased, $Z_j = 9.28$, $p < .001$, $r_s = .40$ (see Table 5). As noted previously, the increase in magnitude of

Table 4
Weighted Effect Size Estimates (g^U) on Performance Measures as a Function of Date of Study

Date of study	Hits	False alarms	Discrimination accuracy	Response criterion
1970s	0.32	−0.28	1.35	0.39
1980s	0.23	−0.38	0.72	0.11
1990s	0.21	−0.41	0.64	0.32

OWN-RACE BIAS SYNTHESIS

Table 5
*Weighted Effect Size Estimates (Z_r) of Influence of
Racial Attitudes and Interracial Contact as a
Function of Date of Study*

Date of study	Attitudes	Contact
1970s	0.06	−0.01
1980s	0.02	0.19
1990s	−0.08	0.27

effect of contact over the past 3 decades may be due to a cohort effect resulting from increases in the opportunities for interracial contact between groups (Chance & Goldstein, 1996). Alternatively, the increase may be due to improved precision and validity in the measures used to assess interracial contact. Nevertheless, it is increasingly evident that the contact hypothesis plays a vital role in our conception of the ORB.

Discussion

The present meta-analysis has empirically reviewed over 30 years of research on the ORB in memory for faces. Thirty-nine research articles were located, involving the combined responses of nearly 5,000 participants. Analyses examined differences in performance on own-race and other-race faces across measures of hit and false alarm responses and across aggregate measures of discrimination accuracy and response criterion. Results of hit and false alarm rates illustrated an ORB mirror-effect pattern in which own-race faces produced a higher proportion of hits and a lower proportion of false alarms compared with other-race faces (see Figure 1). Consistent with this effect, measures of discrimination accuracy demonstrated a significant, moderately sized ORB, accounting for 15% of the variability across samples. Measures of response criterion also showed a significant ORB; however, this effect was considerably smaller, accounting for only 1% of the variability across samples.

In addition, estimates of the influence of both racial attitudes and interracial contact on the ORB were examined across studies. Although no influence of racial attitudes was present in the sample, a small, yet significant, effect of interracial contact was found, accounting for approximately 2% of the variability across the sample. Although racial attitudes appeared to have no direct influence on the ORB, a possible mediating role was indicated by a moderately strong relationship between racial attitudes and interracial contact, accounting for 13% of the variability.

Several study moderators were also examined across the various measures. Results indicated that White participants were more likely to demonstrate the ORB, especially with regard to false alarm responses. Additionally, ORB effects were more likely in measures of discrimination accuracy when presentation and testing were blocked by race of face and when study time was reduced. Measures of response criterion demonstrated ORB effects when stimuli differed between study and test and when the retention interval between study and test was increased. Finally, date of study had a significant influence on both false alarm

and discrimination measures. Results indicated that, over the past 3 decades, the ORB effect appears to have become most prominent in false alarm responses. Measures of the influence of racial attitudes and interracial contact were also affected by date of study, such that the effect of racial attitudes on other-race face recognition has decreased, whereas the effect of interracial contact has increased in more recent years.

Theoretical Implications

The pattern of hit and false alarm responses across studies exhibited a mirror-effect pattern (see Figure 1). This pattern of responses has been demonstrated across a number of manipulations in the literature and has been deemed a "regularity" of recognition memory (Glanzer & Adams, 1985, 1990). It is interesting that this mirror-effect pattern is often captured in aggregate signal detection measures of discrimination accuracy and response criterion, consistent with our meta-analytic results. Much of the debate regarding this phenomenon has involved whether the mirror effect pattern results from a change in the response criterion for each stimulus set, or whether the effects observed on the response criterion measures represent an actual change in the psychological sense of familiarity resulting from the manipulation.

In support of the latter hypothesis, McClelland and Chappell (1998) have proposed a model of recognition memory involving a mechanism of differentiation. As discussed previously, differentiation is a process in which the perceiver focuses attention toward invariant cues that provide the best basis for discriminations within a given stimulus set (Gibson, 1969). McClelland and Chappell model this process by proposing that individuals store features of a given stimulus in memory and that these features (and their associated probabilities) are updated in the representation each time the individual encounters the particular stimulus, thereby resulting in an increase in the psychological sense of familiarity. Furthermore, this increase in the strength of the representation is accompanied by a decrease in the likelihood of responding to a novel, unrelated stimulus. Thus, as McClelland and Chappell conclude, "familiarity breeds differentiation" (p. 726).

McClelland and Chappell's (1998) model effectively simulates the mirror-effect pattern across measures of both discrimination accuracy and response criterion. In doing so, the authors note that the response criterion effects are reproduced despite the fact that the model actually holds the response criterion constant across the stimulus manipulation. Thus, the fluctuation in response criterion is produced as a function of changes in the distributions of new and old items across the familiarity continuum, and not as a result of shifts in the location of the criterion itself. A recent model by Shiffrin and Steyvers (1997) also reproduced these results and was similarly based on the process of differentiation; however, some conceptual differences do exist between the two approaches.

With regard to the ORB, McClelland and Chappell's (1998) model suggests that individuals store own-race faces more accurately and efficiently with respect to the appropriate featural and configural information represented in memory. This accuracy and efficiency may be the result of prior experience (or familiarity) with own-race faces that has led to the ability in attending to the proper invariant aspects of the face. Other-race faces, on the other hand, appear to be encoded in

a less efficient manner, in which fewer or inappropriate cues are selected for storage. When later presented with a recognition task, such differences in encoding result in both differential discrimination accuracy and criterion of responding to own-race versus other-race faces. However, this apparent difference in response criterion occurs as a byproduct of differentiation processes in which the underlying distributions of own- and other-race faces shift along the familiarity continuum. In practical terms, our general familiarity with other-race faces, in the absence of an appropriate representation of features in memory, leads to differential responding in acknowledging the familiarity of the face. As such, this apparent difference in response criterion indicates the role of increased variability in the encoding of featural and/or configural information of other-race faces when compared with the more consistent (less variable) representation of own-race faces.

In summary, the mirror-effect pattern across hit and false alarm responses, together with the associated discrimination accuracy and response criterion effects, suggest a process of differentiation consistent with several recent models of recognition memory (McClelland & Chappell, 1998; Shiffrin & Steyvers, 1997). The implications of this type of model are consistent with the perceptual learning framework outlined previously, including research on the configural–featural (Diamond & Carey, 1986; Rhodes et al., 1989) and race-feature hypotheses (Levin, 1996), as well as the representational model proposed by Valentine and his colleagues (Chiroro & Valentine, 1995; Valentine, 1991). Furthermore, the importance of prior research on the influence of interracial contact, particularly with regard to the effects of discrimination training (e.g., Malpass et al., 1973) and prior experience with other-race faces (Chiroro & Valentine, 1995; Li et al., 1998), are substantiated within this theoretical framework. Although we have previously discussed the potential importance of response criterion measures in the ORB based on findings in our lab (Slone et al., 2000), few studies currently in the literature have documented this effect. Future research that more thoroughly investigates the importance of response criterion can further distinguish its role in the differentiation process.

Applied Considerations

From an applied perspective, several issues merit further discussion. First, the magnitude of the ORB that has been found across many studies, accounting for 15% of the variance in discrimination accuracy, indicates that this is an issue of considerable practical importance. Although our analyses demonstrated that the overall magnitude of the effect on discrimination accuracy has decreased over the past 2 decades, it was also observed that the influence of false alarm responses on the ORB has actually increased during that same period. We believe this to be of great practical significance, as it is precisely the existence of false alarms, namely the erroneous identification of an individual who is not the perpetrator, with which attorneys, judges, and researchers have been most concerned. For example, a recent U.S. Department of Justice report focused on 28 cases in which felony convictions were overturned due to subsequent DNA analyses. In over 85% of those cases, erroneous eyewitness identifications (i.e., false alarms) were the

primary evidence that led to the original conviction (Connors, Lundregan, Miller, & McEwan, 1996).

Second, our moderator analyses indicated that both recognition and lineup identification tasks yield similar ORB estimates across studies. Although a trend was present for lineup tasks to demonstrate a larger ORB effect on correct identifications, more studies involving the use of lineup tasks are needed to better assess the reliability of this effect. Furthermore, as R. C. Lindsay and Wells (1983) noted some time ago, it is important that researchers also manipulate the presence or absence of the target such that they might examine the influence of *diagnosticity* (i.e., the ratio of correct identifications to false identifications) in the other-race lineup situation.

Our moderator analyses also demonstrated that the amount of study time significantly influenced discrimination accuracy in the ORB, particularly through an increase in false alarm responses to other-race faces when study time is limited. Although the application of the laboratory-based term "study time" to the crime situation may seem forced, it should be noted that many crimes involving eyewitnesses occur in a matter of seconds (e.g., assaults, murders, some robberies). This short period of time would involve very limited "study time" for the eyewitness, hence increasing the chances of subsequent false alarms (i.e., mistaken identifications) in cross-race situations.

Moderator analyses also indicated that the length of the retention interval between study and test influenced the ORB through a change in the response criterion. More specifically, this effect indicated that as the length of time increased between study and test, participants increasingly adopted a more liberal response criterion when responding to other-race faces. This liberal response criterion indicated that participants required less evidence from memory (e.g., familiarity or memorability of the face) to respond that they had previously seen an other-race face. In actual cases, the time between viewing the suspect at the crime and later attempting an identification can range between days, weeks, months, and even years. Given this influence of response criterion, the legal community should be cautious of cross-race identifications attempted after such extensive delays.

In the eyewitness literature, researchers have made a distinction between *system variables*—those that are, at least in principle, controllable by the criminal justice system (e.g., interviewing techniques) and *estimator variables*—those that can be manipulated experimentally, but that are not controllable in actual cases; their influence can only be "estimated" post hoc (e.g., the ORB) (Wells, 1978; Wells, Wright, & Bradfield, 1999). Some have suggested that greater research attention should be directed toward system variables, because research results may be more directly applicable to police procedures and legal policy. However, there is one related aspect of the ORB that does involve system variables, namely the procedures used in the construction of identification lineups (Brigham, Meissner, & Wasserman, 1999; Brigham & Ready, 1985). Brigham and Ready (1985) found that race influenced the manner in which individuals constructed lineups, such that both Blacks and Whites used a looser criterion (i.e., more faces were seen as similar to each other, and therefore as useful in a lineup) when constructing lineups of other-race faces as compared with constructing own-race lineups. Hence, there was a tendency to construct fairer lineups (in which the faces were

actually similar to one another) when working with own-race faces. Further research on this application of the ORB seems warranted.

Legal "Safeguards" to the ORB in Eyewitness Identification

As we have discussed elsewhere (Brigham, Wasserman, & Meissner, 1999), several purported "safeguards" are available to defendants accused primarily on the basis of eyewitness evidence, including cross-examination by defense counsel, cautionary instructions to jurors, and expert testimony regarding eyewitness evidence. Although cross-examination has not been shown effective in allowing jurors to distinguish accurate from inaccurate eyewitnesses (R. C. Lindsay, Wells, & O'Connor, 1989; R. C. Lindsay, Wells, & Rumpel, 1981), cautionary jury instructions may have some potential (Cutler, Dexter, & Penrod, 1990; Greene, 1988; Katzev & Wishart, 1985), assuming that they contain accurate information. Unfortunately, such instructions are typically written by legal scholars who have little knowledge of the research findings.

What might more appropriate model jury instructions include? Based on the survey responses of researchers classed as "eyewitness experts" (Kassin et al., 1989) and the results of research meta-analyses, useful model jury instructions could summarize the negative impacts of several factors on the accuracy of eyewitness memory, each of which was listed by over 70% of the experts in the Kassin et al. survey (see also Leippe, 1995). These include short exposure time, high stress, misleading postevent information, and biased lineup instructions. Model instructions could also describe potential problems due to unconscious transference, cross-race identifications, unfair lineups, and the use of showups. They could also point out that expressed confidence or certainty about an identification is not a strong indicator of accuracy (Bothwell, Deffenbacher, & Brigham, 1987; Penrod & Cutler, 1995; Sporer, Penrod, Read, & Cutler, 1995).

The most commonly cited jury instructions are likely those in *United States v. Telfaire* (1972), in which the U.S. Court of Appeals for the District of Columbia endorsed the use of a cautionary instruction on eyewitness evidence. The *Telfaire* instructions state that the juror should evaluate whether the witness "had the capacity and an adequate opportunity to observe the defendant," and whether the witness's identification "was the product of his [sic] own recollection." Jurors are told that they may also take into account "the strength of the identification [certainty]," whether the identification "may have been influenced by the circumstances under which the defendant was presented to him [sic] for identification," and the "length of time that lapsed between the occurrence of the crime and the next opportunity of the witness to see the defendant" (Cutler & Penrod, 1995, pp. 255–256).

Although *Telfaire* was seen by some as a positive step, researchers have noted many shortcomings. For example, the instructions fail to specify in which direction each factor should influence an evaluation of the eyewitness. Furthermore, the *Telfaire* decision was based largely on the five factors listed by the Supreme Court in *Neil v. Biggers* (1972). These factors included (a) the witness's opportunity to view the suspect during the crime, (b) the length of time between the crime and the subsequent identification, (c) the level of certainty demonstrated by the witness during the identification, (d) the (apparent) accuracy of the witness's

MEISSNER AND BRIGHAM

prior description of the suspect, and (e) the witness's degree of attention during the crime. However, only two of these five factors have been clearly supported by research findings (see Brigham, Wasserman, & Meissner, 1999), namely opportunity to view the suspect and the retention interval between viewing and identification of the suspect. Finally, many factors shown by research to be relevant to eyewitness accuracy, such as the ORB, stress, weapon focus, lineup bias, and so forth, are not mentioned in the *Telfaire* instructions.[3]

On a more positive note, the New Jersey Supreme Court recently held that in cases involving a cross-race identification, the defendant is entitled to jury instructions specifically warning jurors about the potential for misidentification of other-race persons (*State v. Cromedy*, 1999). In this case, a Black intruder sexually assaulted a White college student in her apartment. Eight months later the victim saw a man on the street whom she believed to be her assailant. The man was immediately picked up, and the woman identified him 15 min later in a one-person "showup." It is interesting to note that the victim had failed to identify the same man from a photograph lineup only 2 days after the initial assault! The New Jersey Supreme Court, citing in its decision some of the studies included in our meta-analysis, ruled that a cross-race identification, as a subset of eyewitness identification, requires a special jury instruction in the appropriate case. Namely, the instruction should be given when the cross-racial identification is a critical issue in the case, especially when other evidence does not corroborate it. Unfortunately, the instruction advocated by the New Jersey Supreme Court was not directional. The instruction indicated that jurors "may consider, if you think it is appropriate to do so, whether the cross-racial nature of the identification has affected the accuracy of the witness's original perception and/or accuracy of a subsequent identification," without indicating what that effect might be.

With regard to expert testimony, the present meta-analysis results provide additional material that could be presented by an eyewitness expert in cases involving disputed eyewitness evidence. The present findings provide strong evidence of the reliability of the ORB effect, based on the responses of almost 5,000 respondents. The analyses yield meaningful indexes of the strength of the effect, namely that it accounts for 15% of the variance in discrimination accuracy or, alternatively, that participants were over 2.2 times as likely to accurately identify own-race faces as new versus old, when compared with performance on other-race faces. The findings indicate that the majority of errors for other-race faces are false alarms, that is, incorrectly identifying an other-race face as having been seen before. This is the type of error that is generally seen as most harmful in a crime situation. The results show that the ORB is not related to the level of racial prejudice. Finally, factors such as study time and retention interval play an important role in determining when the ORB is most likely to occur.

Given both the reliability of the ORB shown in the present analysis (especially with regard to false alarm responses) and the general agreement among researchers regarding the importance of the phenomenon (Kassin et al., 1989;

---

[3] In his concurring opinion in *Telfaire*, Chief Judge Bazelon urged that juries be warned of the pitfalls of cross-racial identification. Unfortunately, this caution was not included in the final version of the instructions.

Yarmey & Jones, 1983), we advocate the use of expert testimony in cases involving disputed cross-racial eyewitness evidence. Although prior research has demonstrated that the factors influencing eyewitness testimony often reach beyond jurors' common knowledge (Brigham & Bothwell, 1983; Devenport, Penrod, & Cutler, 1997), the courts have often prohibited expert testimony on eyewitness identification, including the ORB (e.g., *People v. Dixon,* 1980; *United States v. Hudson,* 1989; *United States v. Watson,* 1978), ruling that such testimony would not be helpful to jurors. However, in a recent case, *United States v. Norwood* (1996), the U.S. District Court for the District of New Jersey ruled in support of expert testimony on cross-racial identification, along with several other factors. In its decision, the court reasoned that such expert testimony would not confuse or overwhelm the jury. Rather, the "defendant's expert's proposed testimony regarding cross-racial identification was sufficiently tied to facts of [the] case and would be helpful to [the] jury" (p. 1133). The decision relied heavily on the 1985 *United States v. Downing* decision, which held that such expert testimony should (a) properly "fit" the particular features of the case, (b) be based on reliable scientific principles, and (c) not confuse or overwhelm the jury. Expert testimony on cross-racial identifications was also found to be helpful to the jury in *United States v. Stevens* (1984) and *United States v. Smith* (1984).

In closing, the present meta-analysis has yielded many intriguing findings that appear both to bolster our current understanding of the mechanisms responsible for the ORB effect and to illuminate new directions for future research. We believe that previous research has sufficiently underscored the robustness of the phenomenon and illustrated the potential for moderator variables in defining its limits. The current analysis sought only to bring together these findings and to discuss the potential for various theoretical frameworks that might account for the pattern of results across studies. Overall, the ORB was found to be a reliable and generalizable phenomenon, deserving of further theoretical consideration. Moreover, the strong influence of false identifications in the ORB indicates that this issue is of great practical importance as well.

References

References marked with an asterisk indicate studies included in the meta-analysis.

Ahissar, M. & Hochstein, S. (1998). Perceptual learning. In V. Walsh & J. Kulikowski (Eds.), *Perceptual constancy: Why things look as they do* (pp. 455–498). New York: Cambridge University Press.

Allport, G. W. & Kramer, B. M. (1946). Some roots of prejudice. *Journal of Psychology, 22,* 9–39.

Anthony, T., Copper, C. & Mullen, B. (1992). Cross-racial facial identification: A social cognitive integration. *Personality & Social Psychology Bulletin, 18,* 296–301.

*Ayuk, R. E. (1990). Cross-racial identification of transformed, untransformed, and mixed-race faces. *International Journal of Psychology, 25,* 509–527.[A,B,C]

*Barkowitz, P. & Brigham, J. C. (1982). Recognition of faces: Own-race bias, incentive, and time delay. *Journal of Applied Social Psychology, 12,* 255–268.[C,D]

Benton, A. L., Hamsher, K. deS., Varney, N. R. & Spreen, O. (1983). *Contributions to neuropsychological assessment.* New York: Oxford University Press.

*Berger, D. G. (1969). *They all look alike.* Unpublished doctoral dissertation, Vanderbilt University.[A,E,F]

Biederman, I. & Shiffrar, M. M. (1987). Sexing day-old chicks: A case study and expert

systems analysis of a difficult perceptual learning task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13,* 640–645.

Bothwell, R. K., Brigham, J. C. & Malpass, R. S. (1989). Cross-racial identification. *Personality & Social Psychology Bulletin, 15,* 19–25.

Bothwell, R. K., Deffenbacher, K. A. & Brigham, J. C. (1987). Correlation of eyewitness accuracy and confidence: Optimality hypothesis revisited. *Journal of Applied Psychology, 72,* 691–695.

Brigham, J. C. (1981). The accuracy of eyewitness evidence: How do attorneys see it? *Florida Bar Journal, 55,* 714–721.

Brigham, J. C. (1990). Target person distinctiveness and attractiveness as moderator variables in the confidence–accuracy relationship in eyewitness identifications. *Basic & Applied Social Psychology, 11,* 101–115.

*Brigham, J. C. (1993). College students' racial attitudes. *Journal of Applied Social Psychology, 23,* 1933–1967.[G]

*Brigham, J. C. & Barkowitz, P. (1978). Do "They all look alike"?: The effect of race, sex, experience, and attitudes on the ability to recognize faces. *Journal of Applied Social Psychology, 8,* 306–318.[C,E,F,G]

Brigham, J. C. & Bothwell, R. K. (1983). The ability of prospective jurors to estimate the accuracy of eyewitness identifications. *Law & Human Behavior, 7,* 19–30.

*Brigham, J. C., Maass, A., Snyder, L. D. & Spaulding, K. (1982). Accuracy of eyewitness identifications in a field setting. *Journal of Personality and Social Psychology, 42,* 673–681.[A,B,E]

Brigham, J. C. & Malpass, R. S. (1985). The role of experience and contact in the recognition of faces of own- and other-race persons. *Journal of Social Issues, 41(3),* 139–155.

*Brigham, J. C. & Meissner, C. A. (2000, March). Representation and memory for same- and other-race faces. In J. C. Brigham (Chair), *What do we know about the "own-race bias" in face recognition?* Symposium conducted at the biennial meeting of the American Psychology-Law Society, New Orleans, LA.[G]

Brigham, J. C., Meissner, C. A. & Wasserman, A. W. (1999). Applied issues in the construction and expert assessment of photo lineups. *Applied Cognitive Psychology, 13,* S73–S92.

*Brigham, J. C. & Ready, D. J. (1985). Own-race bias in lineup construction. *Law & Human Behavior, 9,* 415–424.[G]

Brigham, J. C., Wasserman, A. W. & Meissner, C. A. (1999). Disputed eyewitness identification evidence: Important legal and scientific issues. *Court Review, 36,* 12–25.

*Brigham, J. C. & Williamson, N. L. (1979). Cross-racial recognition and age: When you're over 60, do they still "all look alike?" *Personality & Social Psychology Bulletin, 5,* 218–222.[C]

Brigham, J. C. & WolfsKeil, M. P. (1983). Opinions of attorneys and law enforcement personnel on the accuracy of eyewitness identifications. *Law & Human Behavior, 7,* 337–349.

Bruce, A. J., Beard, K. W. & Tedford, S. (1997). African Americans' and Caucasian American faces. *Journal of General Psychology, 124,* 143–156.

*Buckhout, R. & Regan, S. (1988). Explorations in research on the other-race effect in face recognition. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory: Current research and issues: Volume 1. Memory in everyday life* (pp. 40–46). New York: Wiley.[A,B]

*Burgess, M. C. R. (1997). *The cross-race effect in facial recognition: A function of expertise?* Unpublished doctoral dissertation, Florida State University.[C]

OWN-RACE BIAS SYNTHESIS

*Byatt, G. & Rhodes, G. (1998). Recognition of own-race and other-race caricatures: Implications for models of face recognition. *Vision Research, 38,* 2455–2468.[B]

*Carroo, A. W. (1986). Other race recognition: A comparison of Black American and African subjects. *Perceptual & Motor Skills, 62,* 135–138.[B]

*Carroo, A. W. (1987). Recognition of faces as a function of race, attitudes, and reported cross-racial friendships. *Perceptual & Motor Skills, 64,* 319–325.[A,B,C,E,F]

Carroo, A. W. (1988). Arousal in differential recognition. *Perceptual & Motor Skills, 67,* 696.

Carter, L. F. (1948). The identification of "racial membership." *Journal of Abnormal & Social Psychology, 51,* 339–341.

Chance, J. E. & Goldstein, A. G. (1979). Reliability of face recognition performance. *Bulletin of the Psychonomic Society, 14,* 115–117.

Chance, J. E. & Goldstein, A. G. (1987). Retention interval and face recognition: Response latency measures. *Bulletin of the Psychonomic Society, 25,* 415–418.

Chance, J. E. & Goldstein, A. G. (1996). The other-race effect and eyewitness identification. In S. L. Sporer, R. S. Malpass, & G. Koehnken (Eds.), *Psychological issues in eyewitness identification* (pp. 153–176). Mahwah, NJ: Erlbaum.

*Chance, J. E., Goldstein, A. G. & McBride, L. (1975). Differential experience and recognition memory for faces. *Journal of Social Psychology, 97,* 243–253.[A]

*Chance, J. E., Turner, A. L. & Goldstein, A. G. (1982). Development of differential recognition for own- and other-race faces. *Journal of Psychology, 112,* 29–37.[A,B,C]

*Chiroro, P. & Valentine, T. (1995). An investigation of the contact hypothesis of the own-race bias in face recognition. *Quarterly Journal of Experimental Psychology, 48A,* 879–894.[A,B,C,E]

Christensen, E. E., Murry, R. C., Holland, K., Reynolds, J., Landay, M. J. & Moore, J. G. (1981). The effect of search time on perception. *Radiology, 138,* 361–365.

Connors, E., Lundregan, T., Miller, N. & McEwan, T. (1996). *Convicted by juries, exonerated by science: Case studies in the use of DNA evidence to establish innocence after trial.* Washington, DC: National Institute of Justice.

*Cross, J. F., Cross, J. & Daly, J. (1971). Sex, race, age, and beauty as factors in recognition of faces. *Perception and Psychophysics, 10,* 393–396.[A,B,E]

Cutler, B. L., Dexter, H. R. & Penrod, S. D. (1990). Nonadversarial methods for sensitizing jurors to eyewitness evidence. *Journal of Applied Social Psychology, 20,* 1197–1207.

Cutler, B. L. & Penrod, S. D. (1995). *Mistaken identification: The eyewitness, psychology, and the law.* Cambridge, England: Cambridge University Press.

Devenport, J. L., Penrod, S. D. & Cutler, B. L. (1997). Eyewitness identification evidence: Evaluating commonsense evaluations. *Psychology, Public Policy, & Law, 3,* 338–361.

*Devine, P. G. & Malpass, R. S. (1985). Orienting strategies in differential face recognition. *Personality and Social Psychology Bulletin, 11,* 33–40.[C]

Diamond, R. & Carey, S. (1986). Why faces are and are not special: An effect of expertise. *Journal of Experimental Psychology: General, 115,* 107–117.

Doty, N. D. (1998). The influence of nationality on the accuracy of face and voice recognition. *American Journal of Psychology, 111,* 191–214.

Elliott, D. N. & Wittenberg, B. H. (1955). Accuracy of identification of Jewish and non-Jewish photographs. *Journal of Abnormal & Social Psychology, 51,* 339–341.

Elliott, E. S., Wills, E. J. & Goldstein, A. G. (1973). The effects of discrimination training on the recognition of White and Oriental faces. *Bulletin of the Psychonomic Society, 2,* 71–73.

Ellis, H. D., Davies, G. M. & McMurran, M. M. (1979). Recall of White and Black faces by White and Black witnesses using the Photofit system. *Human Factors, 21,* 55–59.

MEISSNER AND BRIGHAM

*Ellis, H. D. & Deregowski, J. B. (1981). Within-race and between-race recognition of transformed and untransformed faces. *American Journal of Psychology, 94,* 27–35.[A,B,C]

Ellis, H. D., Deregowski, J. B. & Shepherd, J. W. (1975). Descriptions of White and Black faces by White and Black subjects. *International Journal of Psychology, 10,* 119–123.

*Fallshore, M. & Schooler, J. W. (1995). Verbal vulnerability of perceptual expertise. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 1608–1623.[A]

Farah, M. J., Wilson, K. D., Drain, M. & Tanaka, J. N. (1998). What is "special" about face perception? *Psychological Review, 105,* 482–498.

Feinman, S. & Entwisle, D. R. (1976). Children's ability to recognize other children's faces. *Child Development, 47,* 506–510.

*Galper, R. E. (1973). "Functional race membership" and recognition of faces. *Perceptual & Motor Skills, 37,* 455–462.[A]

Gibson, E. J. (1969). *Principles of perceptual learning and development.* New York: Appleton-Century-Crofts.

Glanzer, M. & Adams, J. K. (1985). The mirror effect in recognition memory. *Memory & Cognition, 13,* 8–20.

Glanzer, M. & Adams, J. K. (1990). The mirror effect in recognition memory: Data and theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 5–16.

Goldstein, A. G. (1979). Race-related variation of facial features: Anthropometric data I. *Bulletin of the Psychonomic Society, 13,* 187–190.

Goldstein, A. G. & Chance, J. E. (1971). Visual recognition memory for complex configurations. *Perception & Psychophysics, 9,* 237–241.

Goldstein, A. G. & Chance, J. E. (1976). Measuring psychological similarity of faces. *Bulletin of the Psychonomic Society, 7,* 407–408.

Goldstein, A. G. & Chance, J. E. (1978). Judging face similarity in own and other races. *Journal of Psychology, 98,* 185–193.

Goldstein, A. G. & Chance, J. E. (1979). Do "foreign" faces really look alike? *Bulletin of the Psychonomic Society, 13,* 111–113.

*Goldstein, A. G. & Chance, J. E. (1980). Memory for faces and schema theory. *Journal of Psychology, 105,* 47–59.[A,B,C]

Goldstein, A. G. & Chance, J. E. (1985). Effects of training on Japanese faces recognition: Reduction of the other-race effect. *Bulletin of the Psychonomic Society, 23,* 211–214.

Green, D. & Swets, J. (1966). *Signal detection theory and psychophysics.* New York: Wiley.

Greene, E. (1988). Judge's instruction on eyewitness testimony: Evaluation and revision. *Journal of Applied Social Psychology, 18,* 252–276.

Haider, H. & Frensch, P. A. (1996). The role of information reduction in skill acquisition. *Cognitive Psychology, 30,* 304–337.

Haider, H. & Frensch, P. A. (1999). Eye movement during skill acquisition: More evidence for the information-reduction hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 172–190.

Hedges, L. V. (1994). Fixed effects models. In H. Cooper, & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 285–299). New York: Sage.

Hedges, L. V. & Olkin, I. (1985). *Statistical methods for meta-analysis.* Orlando, FL: Academic Press.

Helsen, W. & Pauwels, J. M. (1993). The relationship between expertise and visual information processing in sport. In J. L. Starkes, & F. Allard (Eds.), *Cognitive issues in motor expertise* (pp. 109–134). Amsterdam: North-Holland.

OWN-RACE BIAS SYNTHESIS

Hintzman, D. L. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological Review, 95,* 528–551.

Hirshman, E. (1995). Decision processes in recognition memory: Criterion shifts and the list-strength paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 302–313.

Horowitz, E. L. & Horowitz, R. E. (1938). Development of social attitudes in children. *Sociometry, 1,* 301–338.

Hosie, J. A. & Milne, A. B. (1995). Distinctiveness and memory for unfamiliar faces. In T. Valentine (Ed.), *Cognitive and computational aspects of face recognition* (pp. 95–112). London: Routledge.

*Jalbert, N. L. & Getting, J. (1992). Racial and gender issues in facial recognition. In F. Losel, D. Bender, & T. Bliesener (Eds.), *Psychology and law: International perspectives* (pp. 309–316). Berlin, Germany: Walter de Gruyter.[A,B]

Johnson, B. T., Mullen, B. & Salas, E. (1995). Comparison of three major meta-analytic approaches. *Journal of Applied Psychology, 80,* 94–106.

Johnson, K. E. & Mervis, C. B. (1997). Effects of varying levels of expertise on the basic level of categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 126,* 248–277.

Johnson, K. E. & Mervis, C. B. (1998). Impact of intuitive theories on feature recruitment throughout the continuum of expertise. *Memory & Cognition, 26,* 382–401.

Kamphaus, R. W., Beres, K. A., Kaufman, A. S. & Kaufman, N. L. (1996). The Kaufman Assessment Battery for Children (K–ABC). In C. S. Newmark (Ed.), *Major psychological assessment instruments* (pp. 348–399). Boston, MA: Allyn & Bacon.

Kassin, S. M., Ellsworth, P. C. & Smith, V. L. (1989). The "general acceptance" of psychological research on eyewitness testimony: A survey of the experts. *American Psychologist, 44,* 1089–1098.

Katzev, R. D. & Wishart, S. S. (1985). The impact of judicial commentary concerning eyewitness identifications on jury decision making. *Journal of Criminal Law & Criminology, 76,* 733–745.

*Lavrakas, P. J., Buri, J. R. & Mayzner, M. S. (1976). A perspective on the recognition of other race faces. *Perception & Psychophysics, 20,* 475–481.[F]

Leippe, M. R. (1995). The case for expert testimony about eyewitness memory. *Psychology, Public Policy, and Law, 1,* 909–959.

Lesgold, A., Glaser, R., Rubinson, H., Klopfer, D., Feltovich, P. & Wang, Y. (1988). Expertise in a complex skill: Diagnosing x-ray pictures. In M. Chi, R. Glaser, & M. Farr (Eds.), *The nature of expertise* (pp. 311–342). Hillsdale, NJ: Erlbaum.

*Levin, D. T. (1996). Classifying faces by race: The structure of face categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 1364–1382.[A,B,C]

Levin, D. T. & Lacruz, I. (1999, November). *An alternative to the encoding expertise explanation for the cross-race recognition deficit.* Poster presented at the annual meeting of the Psychonomic Society, Los Angeles, CA.

*Li, J. C., Dunning, D. & Malpass, R. S. (1998, March). *Cross-racial identification among European-Americans: Basketball fandom and the contact hypothesis.* Paper presented at the biennial meeting of the American Psychology–Law Society, Redondo Beach, CA.[A,B,C,E,F]

Lickson, J. (1974). *David Charles: The story of the Quincy Five.* Tallahassee, FL: Mockingbird Press, Inc.

*Lindsay, D. S., Jack, P. C. & Christian, M. A. (1991). Other-race face perception. *Journal of Applied Psychology, 76,* 587–589.[A,E]

Lindsay, R. C. & Wells, G. L. (1983). What do we really know about cross-race eyewitness identification? In S. M. A. Lloyd-Boystock, & B. R. Clifford (Eds.),

MEISSNER AND BRIGHAM

Evaluating witness evidence: Recent psychological research and new perspectives (pp. 219–234). New York: Wiley.

Lindsay, R. C., Wells, G. L. & O'Connor, F. J. (1989). Mock-juror belief of accurate and inaccurate eyewitnesses: A replication and extension. *Law & Human Behavior, 13,* 333–339.

Lindsay, R. C., Wells, G. L. & Rumpel, C. M. (1981). Can people detect eyewitness-identification accuracy within and across situations? *Journal of Applied Psychology, 66,* 79–89.

Lindzey, G. & Rogolsky, S. (1950). Prejudice and identification of minority group membership. *Journal of Abnormal & Social Psychology, 45,* 37–53.

Lipsey, M. W. (1994). Identifying potentially interesting variables and analysis opportunities. In H. Cooper, & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 111–124). New York: Russell Sage Foundation.

Luce, T. S. (1974). The role of experience in inter-racial recognition. *Personality & Social Psychology Bulletin, 1,* 39–41.

Macmillan, N. A. & Creelman, C. D. (1990). Response bias: Characteristics of detection theory, threshold theory, and "nonparametric" indexes. *Psychological Bulletin, 107,* 401–413.

Malina, A. C., Bowers, D. A., Millis, S. R. & Uekert, S. (1998). Internal consistency of the Warrington Recognition Memory Test. *Perceptual & Motor Skills, 86,* 1320–1322.

*Malpass, R. S. (1974, September). *Racial bias in eyewitness identification.* Paper presented at the 82nd Annual Convention of the American Psychological Association, New Orleans, LA.[A]

Malpass, R. S. (1981). Training in face recognition. In G. Davies, H. Ellis, & J. Shepherd (Eds.), *Perceiving and remembering faces* (pp. 271–285). London: Academic Press.

Malpass, R. S. (1988). Psychological differentiation, gaze and face recognition. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory: Current research and issues: Volume 1. Memory in everyday life* (pp. 145–150). New York: Wiley.

Malpass, R. S. (1990). An excursion into utilitarian analyses, with side trips. *Behavioral Science Research, 24,* 1–15.

Malpass, R. S., Corey, D., Parada, M., Chavez, J., Bowles, S. & McQuiston, D. (1998, March). *Reliability of face recognition: Take two.* Paper presented at the biennial meeting of the American Psychology–Law Society, Redondo Beach, CA.

Malpass, R. S., Erskine, D. M. & Vaughn, L. L. (1988, April). *Matching own- and other-race faces.* Paper presented at the annual meeting of the Eastern Psychological Association, Buffalo, NY.

*Malpass, R. S. & Kravitz, J. (1969). Recognition for faces of own- and other-race faces. *Journal of Personality and Social Psychology, 13,* 330–334.[A,B,C]

*Malpass, R. S., Lavigueur, H. & Weldon, D. E. (1973). Verbal and visual training in face recognition. *Perception & Psychophysics, 14,* 283–292.[A,B,C]

McClelland, J. L. & Chappell, M. (1998). Familiarity breeds differentiation: A subjective-likelihood approach to the effects of experience in recognition memory. *Psychological Review, 105,* 724–760.

Meissner, C. A. & Brigham, J. C. (in press). A meta-analysis of the verbal overshadowing effect in face identification. *Applied Cognitive Psychology.*

Myles-Worsley, M., Johnston, W. A. & Simons, M. A. (1988). The influence of expertise on x-ray image processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 553–557.

Neil v. Biggers, 409 U.S. 188 (1972).

OWN-RACE BIAS SYNTHESIS

*Ng, W. & Lindsay, R. C. L. (1994). Cross-race facial recognition: Failure of the contact hypothesis. *Journal of Cross-Cultural Psychology, 25,* 217–232.[A,B,C,D,E]

*O'Toole, A. J., Deffenbacher, K. A., Valentin, D. & Abdi, H. (1994). Structural aspects of face recognition and the other-race effect. *Memory & Cognition, 22,* 208–224.[C]

Padgett, K. (1997). *A categorical investigation of the other-race effect in cross-race facial recognition.* Unpublished honor's thesis, Florida State University.

Penrod, S. D. & Cutler, B. L. (1995). Witness confidence and witness accuracy: Assessing their forensic relation. *Psychology, Public Policy, and Law, 1,* 817–845.

People v. Dixon, 410 N.E. 2d 252 (Ill. App. 1980).

*Platz, S. J. & Hosch, H. M. (1988). Cross-racial/ethnic eyewitness identification: A field study. *Journal of Applied Social Psychology, 18,* 972–984.[A,B,E,F]

Proctor, R. W. & Dutta, A. (1995). Perceptual skill. *Skill acquisition and human performance* (pp. 33–65). London: Sage Publications.

Prospero, M., Corey, D., Malpass, R. S., Parada, M. & Schreiber, N. (1996, February). *Is face recognition an ability? On the reliability of face recognition.* Paper presented at the biennial meeting of the American Psychology–Law Society, Hilton Head, SC.

*Rattner, A., Weimann, G. & Fishman, G. (1990). Cross-ethnic identifications and misidentifications by Israelis. *Sociology & Social Research, 74,* 73–79.[A,B]

Raudenbush, S. W. (1994). Random effects models. In H. Cooper, & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 301–322). New York: Sage.

Reingold, E. M., Charness, N., Pomplun, M. & Stampe, D. M. (in press). Visual span in expert chess players: Evidence from eye movements. *Psychological Science.*

*Rhodes, G., Brake, S., Taylor, K. & Tan, S. (1989). Expertise and configural coding in face recognition. *British Journal of Psychology, 80,* 313–331.[A]

Rosenthal, R. (1994). Parametric measures of effect size. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 111–124). New York: Sage.

Secord, P. F., Bevan, W. & Katz, B. (1956). The Negro stereotype and perceptual accentuation. *Journal of Abnormal & Social Psychology, 53,* 78–83.

Shapiro, P. N. & Penrod, S. (1986). Meta-analysis of facial identification studies. *Psychological Bulletin, 100,* 139–156.

Shea, J. B. & Paull, G. (1996). Capturing expertise in sports. In K. A. Ericsson (Ed.), *The road to excellence* (pp. 321–336). Mahwah, NJ: Erlbaum.

Shepherd, J. W. (1981). Social factors in face recognition. In G. Davies, H. Ellis, & J. Shepherd (Eds.), *Perceiving and remembering faces* (pp. 55–79). London: Academic Press.

Shepherd, J. W. & Deregowski, J. B. (1981). Races and faces: A comparison of the responses of Africans and Europeans to faces of the same and different races. *British Journal of Social Psychology, 20,* 125–133.

*Shepherd, J. W., Deregowski, J. B. & Ellis, H. D. (1974). A cross-cultural study of recognition memory for faces. *International Journal of Psychology, 9,* 205–211.[A,B,C,D]

Shiffrin, R. M. & Steyvers, M. (1997). A model of recognition memory: REM—retrieving effectively from memory. *Psychonomic Bulletin & Review, 4,* 145–166.

*Slone, A. E., Brigham, J. C. & Meissner, C. A. (2000). Social and cognitive factors affecting the own-race bias in Whites. *Basic & Applied Social Psychology, 22,* 71–84.[A,B,C,D,E,F,G]

Soukop, V. M., Bimbela, A. & Schiess, M. C. (1999). Recognition memory for faces: Reliability and validity of the Warrington Recognition Memory Test (RMT) in a neurological sample. *Journal of Clinical Psychology in Medical Settings, 6,* 287–293.

*Sporer, S. L. (1992). *Das Wiedererkennen von Gesichtern* [Recognizing faces]. Weinheim, Germany: Beltz/Psychologie Verlags Union. [Experiments 15, 16, and 17].[A,B,C,D]

*Sporer, S. L. (1999, July). *The own-race bias in Germany: Testing the contact hypothesis with Turks and Germans.* Paper presented at the 4th European Conference on Psychology and Law of the American Psychology–Law Association and the European Psychology–Law Association, Dublin, Ireland.[A,B,C,D]

Sporer, S. L., Penrod, S. D., Read, J. D. & Cutler, B. L. (1995). Choosing, confidence, and accuracy: A meta-analysis of the confidence–accuracy relation in eyewitness identification studies. *Psychological Bulletin, 118,* 315–327.

State v. Cromedy, 158 N.J. 112 (1999).

*Swope, T. M. (1994). *Social experience, illusory correlation and facial recognition ability.* Unpublished master's thesis, Florida State University.[A,B,C,E,F,G]

Tanaka, J. W. & Taylor, M. (1991). Object categories and expertise: Is the basic level in the eye of the beholder? *Cognitive Psychology, 23,* 457–482.

*Teitelbaum, S. & Geiselman, R. E. (1997). Observer mood and cross-racial recognition of faces. *Journal of Cross-Cultural Psychology, 28,* 93–106.[A,B,C]

United States v. Downing, 753 F.2d 1224 (3rd Cir. 1985).

United States v. Hudson, 884 F.2d 1016 (7th Cir. 1989).

United States v. Norwood, 939 F.Supp. 1132 (D. N.J. 1996).

United States v. Smith, 736 F.2d 1103 (6th Cir. 1984).

United States v. Stevens, 736 F.2d 1103 (6th Cir. 1984).

United States v. Telfaire, 469 F.2d 552 (D.C. Cir. 1972).

United States v. Watson, 587 F.2d 365 (7th Cir. 1978).

Valentine, T. (1988). Upside-down faces: A review of the effect of inversion upon face recognition. *British Journal of Psychology, 79,* 471–491.

Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion and race on face recognition. *Quarterly Journal of Experimental Psychology, 43A,* 161–204.

Valentine, T. & Bruce, V. (1986). The effect of race, inversion and encoding activity upon face recognition. *Acta Psychologica, 61,* 259–273.

Valentine, T. & Bruce, V. (1988). Mental rotation of faces. *Memory & Cognition, 16,* 556–566.

*Valentine, T. & Endo, M. (1992). Towards an exemplar model of face processing: The effects of race and distinctiveness. *Quarterly Journal of Experimental Psychology, 44A,* 671–703.[A,B,C]

Walk, R. D. (1978). Perceptual learning. In E. C. Carterette, & M. P. Friedman (Eds.), *Handbook of perception: Vol. IX. Perceptual processing* (pp. 257–297). New York: Academic Press.

Warrington, E. K. (1984). *Recognition Memory Test.* Windsor, Ontario, Canada: NFER-Nelson.

Wechsler Memory Scale: Technical manual (3rd ed.). (1997). San Antonio, TX: Psychological Corporation.

*Weimann, G., Fishman, G. & Rattner, A. (1986). Social distance and misidentification. *International Journal of Comparative Sociology, 27,* 217–225.[A,B]

Wells, G. L. (1978). Applied eyewitness-testimony research: System variables and estimator variables. *Journal of Personality & Social Psychology, 36,* 1546–1557.

Wells, G. L., Wright, E. F. & Bradfield, A. L. (1999). Witnesses to crime: Social and cognitive factors governing the validity of people's reports. In R. Roesch, S. D. Hart, & J. R. P. Ogloff (Eds.), *Psychology and law: The state of the discipline* (pp. 53–87). New York: Kluwer Academic/Plenum.

*Wright, D. B., Boyd, C. E. & Tredoux, C. G. (1999, June). *Own-race bias in South Africa and England.* Paper presented at the biennial meeting of the Society of Applied Research in Memory & Cognition, Boulder, CO.[A,B,C,E]

Yarmey, A. D. (1979). *The psychology of eyewitness testimony.* New York: Free Press.

OWN-RACE BIAS SYNTHESIS

Yarmey, A. D. & Jones, H. P. T. (1983). Is the psychology of eyewitness identification a matter of common sense? In S. Lloyd-Bostock, & B. R. Clifford (Eds.), *Evaluating witness evidence: Recent psychological research and new perspectives* (pp. 13–40). Chichester, England: Wiley.

Yin, R. K. (1969). Looking at upside-down faces. *Journal of Experimental Psychology, 81,* 141–145.

Note. Studies contributed one or more of the following estimates: [A]hits; [B]false alarms; [C]discrimination; [D]response criterion; [E]racial attitudes—accuracy; [F]interracial contact—accuracy; [G]attitudes—contact.

C

*177 BEYOND THE KEN? TESTING JURORS' UNDERSTANDING OF EYEWITNESS RELIABILITY EVIDENCE

Richard S. **Schmechel**, Timothy P. O'Toole, Catharine Easterly, Elizabeth F. Loftus [FNa1]

ABSTRACT: Over the past thirty years, researchers have made substantial strides in understanding the workings and limitations of human memory. However, the application of these scientific advances to eyewitness identifications in the criminal justice system, though increasing, has been limited. Trial judges in most jurisdictions exercise their discretionary powers to exclude expert testimony about the reliability of eyewitness identifications. The most common rationale for excluding eyewitness identification expert witnesses is that their findings are not "beyond the ken" of the average juror.

To empirically test this "beyond the ken" rationale, an independent survey of potential jurors in the District of Columbia was designed to investigate whether jurors understand, as a matter of common sense, what makes some eyewitness identifications more or less reliable than others. The survey results, presented in this article, demonstrate that jurors misunderstand how memory generally works and how particular factors, such as the effects of stress or the use of a weapon, affect the accuracy of eyewitness testimony. In light of these findings, judicial practices of excluding expert testimony on the reliability of eyewitness identifications should be reexamined. Wrongful convictions, of which eyewitness identification error is the leading cause, will inevitably continue to result unless jurors can be better educated about these scientific findings.

CITATION: Richard S. **Schmechel**, Timothy P. O'Toole, Catharine Easterly, and Elizabeth F. Loftus, Beyond the Ken? Testing Jurors' Understanding of Eyewitness Reliability Evidence, 46 Jurimetrics J. 177-214 (2006).

*178 Over 75,000 people a year become criminal defendants on the basis of eyewitness identifications. [FN1] Court officers, law enforcement, and jurors are repeatedly called upon to assess the accuracy of witnesses' claims that they remember having seen the accused committing a criminal offense or other significant details about a crime. Yet, without prior training in assessing eyewitness reliability--or, in the case of jurors, the chance to learn how to make this assessment in the courtroom--these key actors in the criminal justice system have only their common sense intuitions as guides when facing a host of difficult questions about witnesses' claims. These questions include the following: Is an eyewitness more reliable if the witness expresses absolute confidence in her identification in trial testimony? Do people remember faces better or worse when they are under stress? When do people's memories begin to fade? Does it matter if a witness is of a different race from the perpetrator? What is the significance of an eyewitness learning new information about the alleged culprit after the event? Is one kind of police identification procedure better than another?

In the late 1970s pioneering psychologists created a new field of social science research dedicated to the study of such questions about eyewitness reliability. They and their successors built upon and extended core research regarding memory and visual perception that dates from the foundation of psychology as a science in the early twentieth century. After thirty years of scientific investigation, the field has found empirical answers to the above questions and many others.

46 JURIMJ 177
46 Jurimetrics J. 177

Such discoveries frequently are relevant to criminal trials, for example:

    • Accuracy-Confidence Correlation: an eyewitness' stated confidence is not a good predictor of identification accuracy;

    • Stress Effects: highly stressful situations may make an experience seem especially vivid, but such stressors can reduce the ability to recall details about a person's face;

    • Time Estimates: eyewitnesses typically overestimate how long an event took to unfold;

    • Cross-race Bias: eyewitnesses are more accurate at identifying members of their own race than members of other races;

    • Postevent Information: eyewitness testimony about an event often reflects not only what a person actually saw, but also information learned later on that unconsciously becomes part of the memory;

    • Presentation Format: witnesses are more likely to misidentify someone when they view all the suspects simultaneously in a group rather than one at a time, in sequence. [FN2]

*179 Eyewitness reliability research uses methods accepted in all sciences. Researchers form hypotheses based on prevailing theories of human memory and cognition, and these hypotheses are tested--sometimes by archival studies (such as of police records) or surveys but mostly by controlled experiments. Data from the experiments are analyzed to see if the results could be explained by chance. Results are peer reviewed before publishing, and it often takes several confirmatory studies of the same hypothesis for a research result to become generally accepted in the scientific community.

For example, in 2004 researchers at Yale University School of Medicine published the results of an experiment on the effects of high and low levels of stress on eyewitness memory. [FN3] The study tested the popular assumption that people are better at remembering events that are stressful and raise a physiological alarm response. [FN4] The scientists hypothesized, based on anecdotal evidence from returning veterans, that this theory may not be true for high stress situations. [FN5] Working with commanders in charge of armed forces prisoner-of-war training, researchers designed an experiment where over 500 healthy military personnel underwent low stress interrogation exercises (using verbal pressure over a forty-minute period) and high stress interrogation (including both verbal and physical confrontation over the same period). Twenty-four hours later, personnel from both groups were asked if their interrogators were present in a photo array of sixteen pictures presented simultaneously, a "live" lineup of fifteen people, or a sequence of sixteen photos shown one at time. [FN6] If participants identified an individual as their interrogator, they were asked their level of confidence in that determination. [FN7] The experimental results show that almost three-fourths of the high stress group and a fourth of the low-stress group could not identify their interrogators, that a sequential identification procedure produced the most accurate results, and that subjects' stated confidence levels were not reliable indicators of accuracy. This large, realistic study has already spurred further reviews of the negative effects of stress on eyewitness reliability, a finding about which there appears to be general consensus emerging among experts. [FN8]

In fact, relative to other scientific research that enters courtrooms, the lack of controversy in the field of eyewitness identification is remarkable. A 2001 survey of established eyewitness researchers found near unanimity that the above-mentioned findings and many others were reliable and established in scientific literature. [FN9] If eyewitness experts are thought of as researchers who offer fact-finders information that helps educate their thinking about the reliability of a *180 witness' testimony, experts may be analogized to other psychological and mental health experts whose testimony is routinely heard by jurors throughout the country. However, the level of consensus and the quantity of research in the eyewitness identification arena (well over 2,000 studies have been done) [FN10] vastly exceeds the research conducted with respect to most other mental health evidence. If one analogizes further and tries to compare the empirical bases and expert consensus in eyewitness research

with fields sometimes considered to be "junk science," like local narcotics trafficking pattern experts who are admitted in most courtrooms, the reliability of eyewitness experts is even more obvious.

Of course, there are limits to what eyewitness research can do. Typically, eyewitness experts are prepared to testify in court about the extent to which the research literature explains how a particular factor, considered alone or in combination with others, likely would affect the reliability of an identification. [FN11] For example, all other things being equal, a victim's identification of an assailant is likely to be less reliable when a weapon was used in the commission of the crime or when the victim views a simultaneous lineup (compared to a crime without a weapon or a sequential lineup showing suspects one-by-one). Experts do not comment on the reliability of a particular witness' memory.

In short, eyewitness reliability research today is an established body of knowledge. It uses well-accepted methodologies. It is part of the research agenda at major universities throughout the world. It is a subject of thousands of peer-reviewed publications. It has existed for decades. There is nearly unanimous consensus among researchers about the field's core findings, and the conclusions of eyewitness reliability research provide an empirical basis for deciding some of the most difficult and pressing questions in this nation's courts.

I. THE TIMELINESS OF RELIABILITY RESEARCH

Given the well-established status of eyewitness research, one would expect such findings to have made their way into courtrooms around the country. Without access to research findings, jurors, judges, and attorneys have only their common sense intuitions with which to gauge an eyewitness who pauses on the stand, points, and says "that's him." Because such in-court identifications are often the most emotionally moving evidence heard in criminal trials, common sense intuitions usually favor crediting the eyewitness' word. Yet, common sense is often wrong. [FN12]

*181 Prompted by DNA exonerations, in the last five years, executive and legislative branches at all levels of government have begun to reform eyewitness identification procedures used by law enforcement officials so as to minimize the risk of altering or biasing eyewitness' memories.

A major catalyst was a 1999 guide to eyewitness evidence published by the National Institute of Justice [FN13] that was followed by a training manual in 2003. [FN14] This guide, drafted by a multidisciplinary panel of eyewitness experts and law enforcement members, was given to every state and federal agency in the country. It presents best-practice recommendations for how to gather identification evidence (such as via a photo array, a lineup, or interviews) in a manner that is nearly as rigorous as rules governing the collection and handling of physical evidence.

Numerous police agencies have adopted new eyewitness procedures in Boston, Minneapolis, and other cities. [FN15] In New Jersey and North Carolina, procedural reforms have been undertaken by the state-level executive branch. Illinois, Virginia, and thirteen other states have completed or pending legislative efforts to mandate state-level reform. [FN16]

The judicial system, too, has become somewhat more receptive to expert testimony and the use of social science in trials. The Federal Rules of Evidence (FRE) were adopted in 1975, immediately preceding the explosion of research on eyewitness reliability. They, and parallel state rules of evidence, [FN17] directly address the admissibility of scientific research by qualified experts in three provisions. Most importantly, FRE 702 and similar rules provide that:

> If scientific, technical, or other specialized knowledge will assist the trier of fact to understand the evidence or to determine a fact in issue, a witness qualified as *182 an expert by knowledge, skill, experience, training, or education, may testify thereto in the form of an opinion or otherwise, if (1) the testimony is based upon sufficient facts or data, (2) the testimony is the product of reliable principles and methods, and (3) the witness has applied the principles and methods reliably to the facts of the case. [FN18]

Other key rules include Rule 704 (allowing expert testimony even when it touches on an ultimate issue to be decided by the fact finder) and Rule 403 (allowing judges to reject any testimony when its probity is "substantially outweighed" by unfair

46 JURIMJ 177
46 Jurimetrics J. 177

prejudice or confusion of the issues).

More than mere codifications of existing practices and case law, FRE 702, 704, and 403 were liberalizing reforms intended to encourage the use of expert witnesses. As expected, overall use of expert witnesses has greatly increased under the rules. [FN19] The Supreme Court's 1995 *Daubert* ruling has opened the door even wider to expert testimony by letting judges apply evidentiary rules even when a field lacks general consensus among experts. Precisely how these rules of evidence have been applied to judicial decisions about eyewitness experts is further explained below. However, the broad trend to increase the use of scientific evidence in courts bears emphasis--such evidence is to be admitted if it is scientifically sound, helpful to laypersons, and not substantially outweighed by any prejudice or confusion.

Another legal trend supports admission of expert testimony regarding eyewitness identification. Over the last thirty years, under the due process clause and the Sixth Amendment, courts around the country have reaffirmed and strengthened the right of an accused to present a complete and zealous defense, including presentation of expert testimony. In *Chambers v. Mississippi*, [FN20] the Supreme Court held that a defendant's right to offer witness testimony on his behalf is one of the most fundamental constitutional rights. [FN21] Building upon this foundation, the Supreme Court, in cases like *Ake v. Oklahoma* [FN22] and *Crane v. *183 Kentucky*, [FN23] has emphasized that the adversary system depends on defendants' ability to contest prosecution theories with expert testimony and to challenge the circumstances under which government evidence was gathered. *Ake* held that state funding of expert psychiatric testimony for defendants was constitutionally required whenever sanity is a significant factor at trial because lay jurors would run the risk of an inaccurate weighing of the facts without expert advice. [FN24] *Crane* held that where an inculpatory confession was nearly the only evidence against the defendant, testimony must be allowed regarding details of the physical and psychological environment present during a confession because these go to the reliability and credibility of the evidence. [FN25] Such cases on the right to present defense witnesses dovetail neatly with the recent *Apprendi v. New Jersey* [FN26] decision that emphasizes the unique role of jurors, rather than judges, as the ultimate fact finders. The overall message is that due process requires that jurors be allowed to consider and resolve contested factual matters, and--if those factual matters involve questions in which expert testimony can be helpful--to hear expert testimony when offered by an accused in a criminal trial.

The powerful stories of wrongful convictions based on flawed eyewitness evidence, the efforts of other government branches and other organizations to reform eyewitness procedures, and broader judicial trends to liberalize evidence rules all combine to make the introduction of eyewitness research into the courts especially timely. Indeed, because these trends all point in the same direction, toward introduction of eyewitness science into the courtroom, one would expect that an eyewitness expert "revolution" would have taken place over the past thirty years. Such an expectation, however, would be misguided.

II. JUDICIAL RESISTANCE TO RELIABILITY RESEARCH

All but a few courts have chosen to ignore clear evidence that denial of defense requests to present expert testimony about eyewitness reliability leads to wrongful convictions. A poignant example of this phenomenon is the Maryland case of Kirk Bloodsworth, [FN27] the namesake of a section of Congress's Innocence Protection Act that provides for postconviction DNA testing. [FN28] Mr. Bloodsworth was twice convicted of the murder and rape of a young girl, chiefly on the basis of one young eyewitness' composite sketch and two eyewitnesses' lineup and in-court identifications. At trial, counsel for Bloodsworth requested that jurors be allowed to hear the findings of an eyewitness expert. Defense counsel's motion *184 was denied by the trial court on grounds that an expert would not be helpful, [FN29] and Bloodsworth then spent nine years in prison before DNA evidence exonerated him. [FN30]

Such resistance is all the more surprising because long before DNA exoneration could prove what is now known--that mistaken eyewitness identifications are the leading cause of wrongful convictions--courts recognized that eyewitness evidence is not as reliable as many jurors think it is. The Supreme Court, first in the 1927 case of Sacco and Vanzetti [FN31] and again in the 1967 case of *United States v. Wade*, [FN32] emphasized that eyewitness identification testimony can be deeply problematic if not carefully monitored. [FN33]

Despite the judiciary's general awareness that eyewitness identifications are problematic, most courts are hostile to the introduction of eyewitness social science in criminal trials. Perhaps this has something to do with the judiciary's own limited understanding of the social science research. One of the largest studies to date on nonexperts' knowledge about eyewitness reliability has shown that individual judges' knowledge of the factors indicating reliable eyewitness claims is flawed. [FN34] In that survey more than two-thirds of respondent judges gave answers contrary to established research in three of the six questions asked about basic issues such as the link between eyewitness confidence and accuracy. [FN35] It may be the very familiarity of judges with a few of the problems with eyewitness testimony that leads many to wrongly assume they need not learn more and, more importantly, that jurors need not learn more either. Such assumptions result in courts around the country refusing to allow criminal defendants to provide jurors with information that could help place eyewitness testimony into a more informed context.

*185 A. Current Case Law

Appellate courts' reluctance to admit eyewitness research usually shows itself in the practice of routinely upholding lower court denials of defendants' motions to allow expert researchers to testify. [FN36] In reaching their decisions, appellate courts have used four main approaches. [FN37]

1. *Pure Discretion Regarding Experts*

The majority rule in approximately twenty-eight state and federal appellate courts is one of pure deference: appellate courts simply say they will defer to trial courts' discretionary decisions. [FN38] The decisions of these majority rule courts offer no standing rules to lower courts as to what fact patterns require eyewitness experts to be admitted. A few such courts, without more, require trial judges to examine the unique facts of the case when eyewitness identifications are key to ***186** the case--a kind of minimal rule that discretion should be exercised on the facts, but one free of any guidance as to how facts should be weighed. [FN39] This "pure discretionary" standard is typically announced by appellate courts with little explanation and often seems to be hollow. For instance, some jurisdictions have said they apply a discretionary standard while at the same time categorically stating that juries do not need expert testimony. [FN40]

With such deferential standards of review, trial court decisions in majority rule jurisdictions are *de facto* final. Review for "abuse of discretion" is vacuous. Since most cases coming up on appeal are denials of defense motions to admit experts, most appellate case law in these "purely discretionary" jurisdictions simply recite trial-level rationales about why experts are excluded and so appear to discourage the admission of eyewitness reliability experts. The appearance that this expert testimony is generally inadmissible is enhanced by the practical barriers to appellate review of any trial court decisions permitting expert testimony. Because the prosecution cannot appeal from an acquittal, cannot generally raise an interlocutory challenge, and has no reason to raise such claims in an appeal from a conviction, a trial judge searching through appellate decisions on the subject will likely look in vain for decisions affirming a decision to admit expert testimony. This is not because such rulings never occur, but because such decisions never get challenged on appeal and published.

Against this majority rule stand state and federal jurisdictions that have circumscribed trial court discretion somewhat or completely.

2. *Limited Discretion to Deny Experts*

Of the twenty-two jurisdictions opposing the pure discretion approach, about eleven purport to require the admission of eyewitness identification experts in limited circumstances, generally where an identification by a stranger is uncorroborated and occurs many weeks after the initial incident. [FN41] Early rulings in Arizona, California, and the Third Circuit started this so-called "modern trend" toward requiring the admission of eyewitness experts when the entire case turns ***187** upon eyewitness evidence. [FN42] However, the "modern trend" has only been followed by a few of the more recent appellate decisions. In fact, Arizona retreated from its broad language in 1989, [FN43] and only one state supreme court followed suit in the 1990s. [FN44]

3. *Limited Discretion to Admit Experts*

A third approach, also in opposition to the pure discretion trend, makes the exclusion of eyewitness experts a default rule except in special circumstances. Approximately six courts have precedent where the denial of experts is required except in circumstances similar to those where admission is required in other minority jurisdictions, such as no corroboration and a significant passage of time between sighting and identification. [FN45] On those unique facts, admission is discretionary.

4. *Prohibition On Experts*

The fourth approach is not to allow eyewitness experts to testify under any circumstance. Until the 1980s, few appellate jurisdictions had given separate attention to whether eyewitness experts could be admitted to testify and a flat prohibition on such testimony was the norm among trial courts. Prohibition of eyewitness expert testimony was largely abandoned in the 1990s when a wave of almost two dozen of the highest state and federal appellate courts adopted the above-mentioned "pure discretion" standard deferring to trial courts. [FN46] However, five high courts, including the Eleventh Circuit, still have a rule of total exclusion. [FN47]

In sum, very few jurisdictions encourage lower courts to provide jurors with eyewitness identification research when sought by criminal defendants. At best, a few jurisdictions require the admissibility of this evidence when its relevance and importance is indisputable. Most jurisdictions do not do even this much; the "pure discretionary" standard used by many courts operates to discourage judges from admitting eyewitness expert testimony, while other courts flatly prohibit the *188 use of such testimony altogether. [FN48] Although the appellate case law has arguably moved in the direction of admissibility over the past thirty years, the movement has been slow and the current standards have the effect of keeping this critical information away from jurors.

B. Rationales for Current Case Law

Why have courts moved so grudgingly in the area of eyewitness experts when the general trend in nearly all other scientific areas has been toward greater admissibility--particularly when evidence is offered by an accused in his own defense? Typically, appellate opinions discussing the admissibility of eyewitness expert testimony provide two reasons: (1) the reliability of the identification is not deemed a critical issue because of corroborating evidence; and (2) the proffered testimony does not pertain to a subject that courts believe is "beyond the ken" of the average juror and therefore "helpful" to the trier of fact. [FN49] Often courts rely on both rationales in explaining why it was permissible to keep the jury from learning about the eyewitness identification research.

1. *The Corroborating Evidence Rationale*

The first of these rationales--that expert testimony about reliability factors is unnecessary because the particular eyewitness identification at issue was not critical to the outcome or was sufficiently corroborated--is a fallacious standard for trial or appellate courts.

A primary problem with this argument is that judicial reliance on corroborating evidence seems to rest on a false premise--namely, that, where other corroborating evidence of guilt exists, defense testimony attacking the reliability of a government eyewitness is of such little importance as to be "irrelevant." If that were true, the proffered eyewitness testimony itself should have been excluded as irrelevant. This obviously never happens because eyewitness testimony is always important evidence. [FN50]

The opportunity to defend oneself before a jury against criminal charges is a bedrock foundation of our adversarial system. A regime where a trial judge decides whether a defense response to prosecution evidence is "important" enough to allow jurors to hear it resembles the Confrontation Clause regime recently condemned by the Supreme Court in *Crawford v. Washington.*

[FN51] In *Crawford*, the question was whether the Sixth Amendment would permit judges to apportion confrontation rights by first deciding whether prosecution evidence *189 was "reliable" enough to make confrontation unnecessary. In rejecting this form of analysis, the Court explained that "dispensing with confrontation because testimony is obviously reliable is akin to dispensing with jury trial because a defendant is obviously guilty. This is not what the Sixth Amendment prescribes." [FN52] Similar language could be used to describe a regime in which the judge, and not the defendant, decides whether prosecution evidence is sufficiently corroborated to render any defense through the presentation of expert testimony unnecessary. Even if there is an independent DNA match identifying the defendant, it simply is not the proper role of the trial judge to exclude proffered expert testimony on government eyewitnesses.

But even if our adversarial system allowed judges to take issues away from juries, looking to corroborating evidence to determine whether an eyewitness identification expert should be permitted is an unsound way to identify those cases where testimony could be significant. First, trial courts deciding admissibility questions in a pretrial posture are poorly situated to resolve questions about whether the "corroborating" evidence is so substantial as to effectively eliminate any potential importance of the expert's proffered testimony to the fact finder. The danger is that prior to hearing all the evidence and without a complete understanding of the case, the trial judge will be swayed by the persuasive power of an eyewitness identification when she views the constellation of facts in a case. The result may be the creation of corroborating evidence out of nothing. [FN53]

Second, particularly in a pretrial posture, it is difficult for a trial court to determine whether corroborating evidence arose independently of the eyewitness identification. In a criminal investigation, evidence emerges from a dynamic context in which each item is affected by the establishment of other evidence. Knowledge of one eyewitness' identification can raise a second person's confidence in their identification or lead police to use more suggestive interrogation tactics with other potential witnesses. [FN54]

Reliance on corroborating evidence, thus, improperly prevents a defendant from responding to important government proof. Because any assessment of corroborating evidence must be made before all the evidence is in, trial courts may prohibit a defense response to prosecution evidence that jurors find to be critical.

*190 Ending reliance on corroborating evidence will admittedly lead to the admission of expert testimony in many more cases. But this is a natural product of the adversarial system in which juries and not judges should resolve disputed questions about the reliability of a piece of government proof. Courts need not fear that such a regime would necessarily invalidate convictions in cases where judges erroneously exclude expert testimony. Harmless error review by appellate courts can effectively ensure that convictions will survive in those rare cases where any expert testimony would truly have been unimportant to the jury's verdict. This appears to be the intent of some trial and appellate courts when applying a corroboration rationale to exclude experts in cases where the evidence is in fact overwhelming; the courts are effectively saying that any exclusion of the expert testimony is harmless because there was so much independent evidence supporting the guilt of the defendant.

The harmless error rule has become well established through many years of application by appellate courts. Appellate courts applying a harmless error standard have a distinct advantage over trial courts in evaluating the significance of corroborating evidence because they have access to more complete information posttrial and because they have clear standards to guide their review. Harmless error standards, moreover, are strongly weighted to ensure that an error will not be deemed "harmless" unless the government can convince the court that the error could not have played any role in the verdict, thus reducing the likelihood that defense expert testimony will be excluded just because a judge, as opposed to a jury, has some intuition that the identification is correct.

Finally, the use of a harmless error standard has one other substantial benefit over a standard blankly excluding experts where there is thought to be corroborating evidence. Harmless error review begins with the determination that a ruling of the trial court, such as the improper exclusion of an expert response to eyewitness testimony, was in fact "error," and then proceeds to determine whether that error mattered in the context of a particular case. This form of analysis ensures that, in the next case, a trial court will permit the expert response because an appellate court has already held that the failure to allow the testimony was error, thus eliminating any sort of speculation about whether the expert testimony would have affected the verdict.

By contrast, where corroborating evidence is cited by appellate courts as the rationale for upholding the exclusion of an eyewitness expert without distinct error and harm analyses, trial courts receive the signal that it is up to them to assess how important the evidence is and whether it will matter to the jury when deciding whether to exclude expert testimony. [FN55]

In sum, whether the evidence corroborating an eyewitness identification is strong or weak, rarely can a reasonable basis be found to disallow expert testimony about the reliability of that identification on grounds that it is merely "irrelevant" or "sufficiently corroborated."

*191 2. *The "Beyond the Ken" Rationale*

The second rationale proffered by courts in current case law--that eyewitness expert testimony is "not beyond the ken" of jurors--is slightly more defensible on its face because it at least applies the proper standard for the admissibility of a particular item of expert testimony. The "beyond the ken" rationale derives from the "helpfulness standard" of Rule 702. The rule requires that the expert testimony "will assist the trier of fact to understand the evidence." [FN56] Even evidence that only marginally improves jurors' ability to determine issues meets this standard. Because of their short opinions, it is not always clear whether courts that refer to the helpfulness standard mean to say that, given corroborating evidence in the case, experts would add nothing more to jurors' ability to judge eyewitness reliability, or whether judges simply believe that jurors know the findings of eyewitness experts as a matter of common sense. In either case, courts say eyewitness research is "not beyond the ken" of jurors.

The problem with this analysis, however, is that (a) it is not apparent from their decisions that trial or appellate judges fully understand what the proffered eyewitness expert testimony would have been; and (b) appellate and trial judges speculate about what jurors understand about eyewitness expert testimony without any hard evidence.

There is considerable evidence that courts do not adequately understand the very expert research they deem to be of no assistance to jurors. Most court opinions, especially the majority following the "pure discretion" approach, contain no literature review and do not parse out which particular aspects of the proffered expert testimony they deem to be common knowledge. In these cases, one can only conclude judges are relying on their experience and intuitions, which, as Wise and Safer demonstrated, are wrong half the time. [FN57] For example, without any expert guidance as to what the research says, only 31% of surveyed judges believed a description of the "forgetting curve" to be true, a research finding that memory often deteriorates quickly shortly after an event, then forgetting becomes more gradual. [FN58] Nonetheless, expert testimony about the "forgetting curve" is typically excluded as common knowledge when proffered by defendants. [FN59] Of course, a few court opinions, particularly those following the modern trend limiting discretion to deny experts, do take care to address particular eyewitness research findings. However, even these cases rarely cite to research and end up with unscientific and contrary holdings, disagreeing, for example, on whether the lack of significant correlation between a witness' confidence and their accuracy is commonly known. [FN60] Stress, too, and its *192 debilitating effects on memory, is thought to be common sense to some of those few courts who have considered the matter counterintuitive to others. [FN61] In short, while most appellate courts do not address specific eyewitness research findings and do not examine the research, those that do have generated conflicting precedent that gives trial courts no sure guidance in exercising their discretion.

Most importantly, courts also do not understand what jurors know about eyewitness reliability. Until now, no one has appropriately measured what jurors know about eyewitness reliability. Research into laypersons' knowledge of eyewitness reliability factors has been sparse. Eyewitness researchers, in the course of conducting more targeted experiments, have repeatedly found their subjects to be unaware of basic scientific findings of the field. There also have been some small, non-randomized surveys about juror understanding of eyewitness reliability that indicated a general lack of knowledge. [FN62] Based on these small surveys and their laboratory experience, scientists studying eyewitness reliability have come to believe that ordinary lay jurors know little about the subject as a matter of common sense.

These researchers' opinions are in sharp contrast to judicial opinions. For example, recent surveys have shown that

mug-shot-induced bias--the fact that prior exposure to mugshots of a suspect increases the likelihood that a witness will later choose the suspect from a lineup--was thought to be a matter of common sense by only 13% of researchers compared with 74% of judges. [FN63] The rift between the courts' and researchers' beliefs about what jurors understand about eyewitness reliability has been unresolved because no one has performed a large, empirical study of jurors' knowledge. Thus, judicial rulings about what is "beyond the ken" of jurors are, at best, speculation based solely on judges' experience and intuition. [FN64]

Neither judicial rationale for excluding eyewitness experts--that their testimony is unnecessary because of corroborating evidence or not helpful because it is not beyond the ken of jurors--holds up under examination. The first rationale is logically and legally flawed, and the second is mere speculation.

To better ground their assertions about what is "beyond the ken," judges would have to understand both what eyewitness research says and whether *193 laypersons' common sense understanding of these issues parallels the research findings. If jurors can be shown to understand eyewitness research findings as a matter of common sense, then the judicial rationale for excluding experts as "unhelpful" is valid. But, if eyewitness research findings are not known to jurors, the only plausible rationale for excluding eyewitness experts is wrong and courts should reevaluate their positions on eyewitness expert admissibility.

III. EMPIRICAL FINDINGS

In the winter of 2004, lawyers from the Public Defender Service for the District of Columbia (PDS) began investigating whether jurors do understand as a matter of common sense what makes some eyewitness identifications more or less reliable than others. [FN65] Preliminary investigation showed that despite extensive academic research into the reliability of eyewitnesses, there was little direct research on the extent to which nonexperts' intuitions match experts' findings. [FN66] The surveys indicate that while many attorneys have sound common-sense opinions about factors like cross-racial identification, the effects of stress or violence, and the importance of unbiased lineup instructions, most lack an accurate understanding of eyewitness confidence or bias in simultaneous (versus sequential) lineups. [FN67] Surveys of judges (mentioned above) indicate they have correct intuitions at roughly the same rates as attorneys. [FN68] However, there have been only a handful of surveys about the knowledge of nonexperts outside the legal system, and most of these have been nonrandomized studies using under-graduate or law students. [FN69] These latter surveys have used a diagnostic test called the Knowledge of Eyewitness Behavior Questionnaire (KEBQ) that poses scenarios with multiple factors relevant to eyewitness identifications. As with the judicial and attorney surveys, a significant portion of students taking the KEBQ were found to have some awareness of factors, such as cross-racial identification, but grossly misunderstand others, like the fact that training does not improve a witness' accuracy. [FN70]

*194 There has been significantly more "indirect" research into the attitudes of nonexperts about eyewitness reliability, such as research that does not attempt to survey a population's self-described understanding of reliability but instead relies upon smaller studies of how laypersons reason about eyewitness identification factors. [FN71] For example, postdiction studies required respondents to read summaries of eyewitness identification experiments and then predict the rate of accurate identifications in the experiment. Similarly, judgment studies put nonexperts through mock trials and assessed jurors' sensitivity to different aspects of the case through questionnaires and their final verdict. Considered together, the quantity of this "indirect" research into laypersons' knowledge of eyewitness research findings suggests jurors' common sense reasoning on these matters is often flawed. However, this kind of research lacks the persuasive value of direct surveys assessing the knowledge of the jury pool. [FN72]

To describe the Washington, D.C. juror pool, PDS decided to conduct its own research on the issue. [FN73] PDS accordingly commissioned a Washington, D.C. polling firm to survey approximately 1,000 potential District of Columbia jurors. The survey results, discussed in detail below, demonstrate that judicial assertions about jurors' ability to appraise eyewitness identifications are wrong. Jurors suffer from a basic misunderstanding of how memory generally works and do not understand how particular factors, such as the effects of stress or the use of a weapon, affect the accuracy of eyewitness testimony.

46 JURIMJ 177
46 Jurimetrics J. 177

A. Mechanics of the PDS Survey

In the Fall of 2003 and early Winter of 2004, Peter D. Hart Research Associates, Inc. (Hart Research) worked together with attorneys at PDS and Dr. Elizabeth Loftus to prepare a survey that would discern what potential District of Columbia jurors understand about memory in general and the reliability of eyewitness identification evidence in particular. The questionnaire consisted of approximately twenty questions about whether jurors believe eyewitness testimony is generally reliable and also about the specific factors jurors believe would make an eyewitness identification more or less reliable. [FN74]

*195 Between February 18, 2004, and February 23, 2004, Hart Research conducted a telephone survey in the District of Columbia. Residential phone numbers were chosen at random from the District of Columbia area code. One thousand two hundred ninety-six potential jurors were identified as U.S. citizens who were at least eighteen years of age and not currently on probation or parole. Participants were also asked demographic questions about education, neighborhood, age, employment, party affiliation, prior jury service, language, race, income and exposure to the criminal justice system. Ultimately, 1,007 potential jurors completed the survey. The margin of error (95% confidence) for the survey was plus-or-minus 3.1 percentage points.

1. Juror Misunderstandings of Memory in General

Human memory does not record events like a video recorder. [FN75] In the first place, human memory is more selective than a video camera. The sensory environment contains a vast amount of information, but the memory process perceives and accurately records only a very small percentage of that information. Second, because the act of remembering is reconstructive, [FN76] akin to putting puzzle pieces together, human memory can change in dramatic and unexpected ways because of the passage of time or subsequent events, such as exposure to "post-event" information like conversations with other witnesses or media reports. Third, memory can also be altered through the reconstruction process. Questioning a witness about what he or she perceived and requiring the witness to reconstruct the experience can cause the witness' memory to change by unconsciously blending the actual fragments of memory of the event with information provided during the memory retrieval process.

These characteristics of human memory have profound implications with regard to the accuracy of eyewitness claims that they "remember" seeing the accused or other key details about the crime. For a memory to be reliable, the witness must have accurately perceived the event, and the witness' memory must not have degraded over time or been polluted by postevent information and questioning. But more importantly, for jurors to fairly assess whether this claim is accurate, they must understand memory's complexity, selectivity, and malleability. Jurors must also understand what specific factors affect perception and encoding of memories, what factors can pollute memory, and what factors in the re-creation process can distort a witness' "memory" of an event.

PDS's survey of potential jurors in the District of Columbia suggests that juror understanding of these subjects fails at even the most basic level. Several survey questions, for example, were designed to test jurors' general understandings of the workings of human memory. One question asked whether *196 "the act of remembering a traumatic event [was] like a video recording in that one can recall details as if they had been imprinted or burned into one's brain." Over half of the respondents (52%) either thought this statement was true or did not know whether it was true. Indeed, 46% of potential jurors believe that the witness on the stand is effectively narrating a video recording of events that she can see in her "mind's eye" for jurors.

Other results demonstrated similar deficits of knowledge on the most basic level about how memory works. The survey asked potential jurors to assess the reliability of their own memories. Almost two-thirds of the respondents (66%) thought the statement "I never forget a face" applied "very well" or "fairly well" to them. Likewise, more than three out of four respondents (77%) thought that the phrase "I have an excellent memory" applied "very well or "fairly well" to them. The fact that such large majorities of people tend to believe their memories are above average suggests that potential jurors may begin each trial with unwarranted confidence in memory and the ability to identify faces generally. Such confidence may well be transferred to the testifying witnesses and cause jurors to overestimate the accuracy of witness memories as well.

46 JURIMJ 177
46 Jurimetrics J. 177

2. *Specific Reliability Factors--Weapon Focus*

For more than twenty-five years, social scientists have posited that the presence of a weapon during a crime attracts the attention of the witness to the weapon, reducing attention to the culprit's facial and physical characteristics. This phenomenon is often referred to as "weapon focus." [FN77] A series of controlled studies have now validated the existence of this effect. [FN78] For example, experiments have been conducted involving videotaped robberies with some of the culprits brandishing a handgun and others concealing the gun. These studies have repeatedly demonstrated that witness accuracy was better when the gun was concealed than when the gun was brandished.

*197 Some courts have speculated that "weapon focus" is among those features of memory that jurors understand as a "matter of common sense." [FN79] The PDS survey asked potential jurors whether they thought that the fact that "a weapon is involved in the crime" tends to make "an eyewitness' memory about the details of the crime more reliable, less reliable or [would have] no effect." Thirty-seven percent of the respondents actually thought the presence of a weapon would make a witness' memory for event details *more* reliable, while thirty-three percent of the respondents thought that the presence of a weapon either would have no effect or were not sure of what effect a weapon would have. Only three out of ten potential jurors correctly understood that the presence of a weapon tends to make an eyewitness' memory for details less reliable.

3. *Specific Reliability Factors--Presence of Violence or Stress*

Social science studies have also shown that a person's ability to recall details of an event is likely worse if a witness has observed a violent event as opposed to a nonviolent one. [FN80] This is particularly true for the peripheral details.

To find out whether jurors understand the effect that violence has on witness identifications, the PDS survey asked potential jurors whether they thought that the fact that "a crime is violent" tends to make "an eyewitness' memory about the details of the crime more reliable, less reliable or [would have] no effect." Thirty-nine percent of the respondents thought that event violence would make a witness' memory for event details more reliable, [FN81] while 33% of the respondents thought that event violence either would have no effect or were not sure of what effect event violence would have. Only three out of ten potential jurors correctly understood that event violence tends to make an eyewitness' memory for details less reliable.

4. *Specific Reliability Factors--Duration of the Incident*

For over a century, social scientists have been conducting experiments concerning people's ability to estimate the duration of a stressful incident. Those studies have consistently shown not only that most people have enormous difficulty estimating the length of these events but also that the vast majority of these errors are in the same direction--toward overestimating the duration of a *198 stressful event. Overestimation can vary substantially depending upon the amount of stress accompanying the event. [FN82]

Thus, the longer an eyewitness has to observe a particular face, the more accurate her identification becomes. Although jurors understand this principle as a matter of common sense, they do not also understand the unreliability of witnesses' subjective time estimates. Often, the only piece of information a juror has about the length of exposure time is the witness' estimate. This estimate exaggerates the exposure time.

The PDS survey shows that jurors do not understand this phenomenon as a matter of common sense. Over 40% of survey respondents either thought that witness time estimates were accurate or were not sure whether such estimates were accurate. While that alone indicates that a significant number of jurors would be likely to overestimate exposure times and thus witness reliability, the beliefs of those who are more skeptical of witness time estimates are even more troubling. Of those who correctly understood that witnesses themselves are not good at evaluating how long an event took to unfold, a sizeable portion (about 25%) believed that witnesses *underestimate* the actual time. In all, 63% of the survey respondents do not understand what scientific research has demonstrated about a witness' ability to gauge the duration of an event. The jurors either believed witnesses' subjective time estimates or thought that witnesses tended to actually see a face for longer than claimed. Only 37% of

46 JURIMJ 177
46 Jurimetrics J. 177

the total respondents correctly understood events unfold faster than witnesses think they do.

5. *Specific Reliability Factors--Confidence*

With regularity, a witness in court will express 100% confidence in an identification. But is the identification of a "confident" witness more reliable than the identification of a less certain one? Witnesses who are highly confident in their identifications are only slightly more likely to be correct as compared to witnesses who are less sure of the identifications. In other words, the correlation between confidence and accuracy is weak. [FN83]

Eyewitness accuracy is a complex product of many factors. Factors such as the witness' eyesight and concentration, the amount of lighting, the length of *199 exposure, the quality of the view, whether a weapon or violence was involved, and the procedures used by police to obtain the identification all help to determine whether a particular identification is an accurate one.

By contrast, confidence is a product of personality and social factors of which accuracy of observation is only a minor part. A witness' confidence will also depend on how self-confident the witness is to begin with and what interactions the witness has had with others to boost or undermine that confidence. For example, studies have shown that confidence is highly malleable and can be substantially increased by many postevent factors, including confirming feedback. [FN84] Witness confidence can accordingly increase after the incident through the use of procedures that do not in any way enhance the accuracy of the original identification and may undermine it. The weak correlation between confidence and accuracy that may have existed immediately after the incident is thus often destroyed after a witness' confidence level is raised or lowered through "contaminating" exposure to feedback.

The PDS survey results demonstrate that jurors do not understand the relationship between confidence and accuracy. For example, one survey question asked respondents to compare the reliability of a witness who was "absolutely certain" of an identification with that of a witness who was not. A plurality of respondents, 31%, found the "absolutely certain" witness to be "much more reliable." Moreover, only 17% of the respondents correctly understood the slight correlation between confidence and accuracy. Thus, a majority of the respondents demonstrate a fundamental misunderstanding about the confidence-accuracy correlation. [FN85]

Another confidence-accuracy survey question confirmed these findings. Nearly 40% of survey respondents agreed that "an eyewitness' level of confidence in his or her identification is an excellent indicator of that eyewitness' reliability." Thus, four out often potential jurors, absent education on this subject, would assess witness testimony under the mistaken impression that there is a very strong correlation between witness confidence and witness accuracy. Moreover, even though 55% of polled jurors correctly reject the notion that confidence is an "excellent indicator" of accuracy, the responses to the earlier question demonstrate that these jurors do not understand whether any confidence-accuracy correlation exists and, if so, what that correlation is. Thus, these survey results also make clear that jurors have no meaningful idea of how to evaluate witness statements of confidence and are likely to substantially overestimate the reliability of a confident witness.

*200 6. *Specific Reliability Factors--Police Officers*

Studies have shown that police officers are sometimes able to provide more detailed accounts of the event and that they are sometimes less susceptible to the effects of postevent misinformation when compared to laypersons. But research has also shown that police officers perform no better at identifying faces than other citizens, [FN86] which means that their eyewitness identifications are generally no more reliable than anyone else's. [FN87]

To test whether jurors understood this fact, the PDS survey asked respondents to compare the reliability of an eyewitness identification by "a police officer" with the reliability of an identification by "an average citizen." Sixty percent of the respondents failed to understand that the two identifications were equally reliable, and many of these respondents (22% of the total) believed police officer testimony to be "much more reliable."

46 JURIMJ 177
46 Jurimetrics J. 177

7. *Specific Reliability Factors--Cross-Racial Impairment*

A variety of studies have shown that eyewitnesses experience a "cross-racial impairment" when identifying members of another race. Eyewitnesses are better at identifying members of their own race and have difficulty identifying members of other races. [FN88]

When asked to compare the reliability of an identification by an eyewitness "of the same race as the person being identified" with the reliability of an identification by an eyewitness "of a different race" from the suspect, almost two-thirds of jurors surveyed indicated that they are ill-informed about the inaccuracy of cross-racial identification: A large plurality of the survey respondents (48%) thought cross-race and same-race identifications are of equal reliability, and many of the other respondents either did not know or thought a cross-racial identification would be more reliable (11%). Only 36% of the survey respondents understand that a cross-racial identification may be less reliable.

8. *Specific Reliability Factors--Show-Ups*

A show-up occurs when police display a single suspect to the witness and ask whether that suspect is the culprit. Both the Supreme Court and the Attorney General have recognized the inherently suggestive nature of show-up *201 procedures, [FN89] and the Supreme Court has expressly acknowledged the greater risk of misidentification involved in show-ups as compared to lineups. [FN90] There is strong empirical evidence that show-ups are more likely to yield false identifications than properly constructed lineups. The risk of generating an unreliable identification by using a show-up identification procedure, moreover, grows substantially more pronounced over time [FN91] especially if the suspect is wearing clothing similar to the culprit's.

The PDS survey indicates substantial confusion among potential jurors about the reliability of show-up procedures. While this was one of the few areas where a majority of potential jurors appear to have intuitions similar to research findings, a substantial minority of jurors still did not understand this concept. A quarter of potential jurors believed that a show-up is either *more* reliable than a lineup procedure or that the two procedures are equally reliable.

9. *Specific Reliability Factors--Lineup Instructions*

A lineup procedure occurs when a witness is shown a group of people and asked to attempt an identification. Research demonstrates that during a lineup procedure eyewitnesses tend to engage in a comparative analysis, that is, they identify the person from the lineup who, in the opinion of the eyewitness, *looks most like* the culprit relative to other members of the lineup. [FN92] Where the actual culprit is not in the lineup, this relative judgment process poses a grave danger of misidentification because there will always be someone who looks more like the culprit than the remaining lineup members.

Research has demonstrated that a witness' willingness to make a relative judgment about the culprit, in a lineup that does not include the culprit, is considerably less when the witness is warned that the culprit may not be present *202 in the lineup. [FN93] Moreover, this instruction results in the same number of accurate identifications when the culprit is actually present in the lineup.

Potential jurors do not appear to understand the role or importance of such instructions. Over half the respondents in the PDS survey (51%) thought that an identification would be *more* reliable if the eyewitness was *not* instructed about the culprit's potential absence, and an additional 21% either thought it did not matter whether such an instruction was given or were not sure. Only 28% of the respondents thought that a lineup would be more reliable if the eyewitness *was* instructed that the actual culprit "may or may not be in the lineup." These numbers were virtually the same for photographic arrays. [FN94] The survey thus demonstrates substantial juror confusion on the importance of lineup and photo array instructions.

10. *Specific Reliability Factors--Sequential Identification*

Another way to avoid the relative-judgment problem is the use of a sequential lineup. Unlike a traditional lineup in which an eyewitness views the lineup members simultaneously, in the sequential lineup the eyewitness views the members one at a time and decides individually whether each person is the culprit before viewing the next member. Studies have shown that this procedure reduces the number of mistaken identifications, particularly when the witness does not know how many members will be viewed and thus cannot anticipate when the process will end. The sequential procedure produces fewer mistaken identifications in lineups that do not contain the actual perpetrator, but it does not significantly impair eyewitnesses' abilities to identify the perpetrator in perpetrator-present lineups. [FN95]

*203 Potential jurors are unaware of the benefits of sequential identifications procedures. When asked to compare the reliability of a procedure in which a "witness views a lineup of potential suspects standing next to one another" with a procedure in which the witness "views potential suspects one at a time," over three-quarters of the respondents (76%) either thought the reliability of a simultaneous live lineup was better than or equal to that of the sequential lineup, or were not sure which process was better. With respect to photo lineups, the numbers are similar--61% of the respondents either thought the reliability of a simultaneous photo lineup was better than or equal to that of a sequential photo lineup or were not sure which process was better. In both cases, therefore, potential jurors seemed not to understand the importance of a sequential lineup in securing an accurate identification.

11. *Specific Reliability Factors--"Double Blind" Procedures*

The behavior of the person who administers the identification procedure can influence the reliability of eyewitness testimony itself. [FN96] If the person who administers the live or photographic lineup is the case detective or some other investigator who knows the identity of the suspect, a substantial danger exists that the person conducting the lineup will communicate the suspect's identity to the witness. [FN97] This can happen consciously or unconsciously, and through both verbal and nonverbal means. Moreover, as noted above, the administrator of the lineup can also cause mistaken eyewitnesses to develop high levels of false certainty by providing postidentification feedback about the supposed accuracy of the identification. All of these problems can be substantially reduced through the use of "double blind" procedures, that is, when lineups are administered by someone who does not know which lineup member is the suspect and which ones are fillers.

With respect to live lineups, a bare majority of potential jurors (55%) appears to have at least some grasp of the importance of conducting "a lineup where the police officer running the lineup is unaware who the suspect is," although a substantial minority of jurors (45%) do not understand the importance of this concept at all. Specifically, one fifth of the jurors incorrectly believe that a live lineup where police know the identity of the suspect is *more* reliable than a "double blind" procedure, and another 27% of respondents believe either that the two procedures are equally reliable or were unsure of the difference between the two procedures. This means that approximately 5 jurors in a panel of 12 will start *204 a trial with a basic misunderstanding about the importance of "double blind" procedures, a misunderstanding that will obviously continue throughout the trial and into jury deliberations in the absence of accurate, authoritative information.

This problem is even more pronounced with photographic arrays. A majority of potential jurors (52%) do not understand the importance of conducting "a photo array where the police officer running the photo array is unaware who the suspect is." Specifically, 30% of the jurors incorrectly believed that a photo array where police know the identity of the suspect is *more* reliable than a "double blind" procedure, and another 22% of respondents believed either that the two procedures are equally reliable or were unsure of the difference.

B. Summary of Survey Findings

The PDS survey documents that a substantial number of jurors come to each trial with basic misunderstandings about the way memory works in general and about specific factors that can affect the reliability of eyewitness identifications. The findings undermine previous judicial speculation about what jurors know as a "matter of common sense." As an empirical matter, the PDS poll shows that significant numbers of jurors (often substantial majorities) do not understand concepts like weapon focus, the effects of stress, the tendency of witnesses to overestimate exposure time, and the lack of meaningful cor-

relation between witnesses' stated confidence and accuracy in making an identification. Jurors also place unwarranted stock in the eyewitness abilities of police officers, they overestimate the reliability of cross-racial identifications, and they have minimal understanding of how police procedures can affect the accuracy of an eyewitness identification.

C. Potential Limitations of Survey Findings

Any survey is limited in its ability to exactly determine the knowledge of laypersons. [FN98] However, if anything, it is likely that the survey methodology used in this instance *overestimates* the juror understanding of eyewitness reliability factors. First, the questions asked often singled out the relevant variables (rather than the juror having to analyze the facts of an identification and say that, for example, the presence of a weapon is significant). Second, there is a base rate of accuracy that comes from survey participants guessing among the multiple-choices, despite being given the option to say they did not know. Therefore, survey participants' correct responses to the survey questions likely overestimate the extent to which a juror would take proper account of that variable in an actual case.

The precise applicability of these survey results to other jurisdictions cannot be determined, though the pool of potential District of Columbia jurors surveyed is likely similar to many other urban jury pools nationwide. The complete results *205 of the survey list the detailed demographic information of survey participants. Notably, the education level of participants was quite high (45% received a bachelors or higher degree) and reflected a wide range of ages and incomes.

Over the past 30 years, researchers have made substantial strides in understanding the workings and limitations of human memory. However, the application of these scientific advances to eyewitness identifications in the criminal justice system, though increasing, has been limited. Trial judges in most jurisdictions routinely exercise their discretionary powers to exclude expert testimony about the reliability of eyewitness identifications. This is so despite changes to the rules of evidence, Supreme Court opinions acknowledging defendants' rights to put on testimony in their defense, and attempts by executive and legislative branches to give greater scrutiny to eyewitness identifications.

The most common rationale offered for excluding eyewitness identification expert witnesses is that their findings are not "beyond the ken" of the average juror. To test this rationale, a survey of potential jurors in the District of Columbia was designed to investigate whether jurors understand, as a matter of common sense, what makes some eyewitness identifications more or less reliable than others. The survey results presented above demonstrate that jurors misunderstand how memory generally works and how particular factors, such as the effects of stress or the use of a weapon, affect the accuracy of eyewitness testimony. Jurors also misunderstand how eyewitnesses' stated levels of certainty correspond with accuracy. These findings demonstrate that there are flaws in judicial intuitions about what jurors do and do not understand.

Judges, prosecutors, and defense attorneys alike need to be educated about the perils of eyewitness identification. The common sense assumptions of these groups may be somewhat better than the average juror by virtue of greater experience with the criminal justice system, but even these groups lack an adequate understanding of eyewitness reliability. Defense attorneys will not introduce experts or file suppression motions, prosecutors will not discern credible allegations, and judges will not hear counsel's arguments on eyewitness reliability--unless they all reach beyond their commonsense intuitions. Tens of thousands of convictions each year are made on the basis of eyewitness evidence about which jurors' common sense is mistaken. To safeguard against wrongful convictions on eyewitness evidence, the justice system must replace false intuitions with scientifically based opinions.

*206 APPENDIX

SURVEY QUESTION AND ANSWER DATA

All results are shown as percentages unless otherwise noted. Correct answers to knowledge-based questions are in bold.

1a. What is the last grade that you completed in school?

Grade school	1
Some high school	7
High school graduate	27
Some college, no degree	13
Vocational training/2-year college	6
4-year college/bachelor's degree	22
Some postgraduate work, no degree	3
2-3 years postgraduate work/master's degree	14
Doctoral/law degree	6
Not sure/refused	1

1b. In which quadrant of Washington, D.C., do you live--Southeast, Southwest, Northeast, or Northwest?

Southeast	22
Southwest	4
Northeast	26
Northwest	48

Not sure/refused   -

---

2. As you may know, to be eligible for jury duty in Washington, D.C., you must be a U.S. citizen, at least eighteen years old, and a resident of the District of Columbia. You must read, speak, and understand English, and you must NOT be on probation or parole. You DO NOT have to be registered to vote. Given these requirements, are you currently qualified for jury duty in Washington, D.C.?

Yes, qualified to serve on a jury   100   CONTINUE

No, not qualified to serve on a jury   -   **TERMINATE**

Not sure   -

---

*207 3. For each of the following phrases, please tell me whether it applies to you very well, fairly well, just somewhat well, or not very well.

	Applies Very Well	Applies Fairly Well	Applies Just Somewhat Well	Does Not Apply Very Well	Not Sure
I never forget a face	38	28	24	9	1
I have an excellent memory	44	33	17	6	-

4a. In some criminal trials, an eyewitness may testify that the person on trial, who is also called the defendant, was the person the eyewitness saw committing the crime. In general, do you consider this sort of eyewitness identification to be a very reliable, somewhat reliable, not very reliable, or unreliable form of evidence?

Very reliable   18

Somewhat reliable   65

Not very reliable   11

Unreliable	4

Not sure	2

      4b. And suppose that an eyewitness has no motivation to lie, and genuinely believes that his or her identification of the defendant is accurate. In those cases, do you consider eyewitness identifications in criminal trials to be very reliable, somewhat reliable, not very reliable, or unreliable forms of evidence?

Very reliable	25

Somewhat reliable	62

Not very reliable	8

Unreliable	3

Not sure	2

*208 5. Now I'm going to describe several pairs of eyewitnesses who might testify in a criminal trial. I'd like you to tell me whether you would consider the first or second eyewitness to be the more reliable ("MR") eyewitness, or whether you would consider both eyewitnesses to be equally reliable ("ER"), or whether neither would be reliable ("N"). **(If necessary ask:)** Would the (first/second) eyewitness be "much more" reliable ("MMR") or only "slightly more" reliable ("SMR")?

The First Eyewitness		The Second Eyewitness		ER	N	Not Sure
MMR	SMR	SMR	MMR			

5a. One eyewitness says they are absolutely certain of their identification of the

criminal defendant, AND another eyewitness does NOT say they are absolutely

certain of their identification of the criminal defendant.

| | 31 | 17 | 12 | 5 | 26 | 4 | 5 |

5b. One eyewitness is a police officer, AND another eyewitness is an average

citizen.

| | 22 | 15 | 7 | 9 | **40** | 3 | 4 |

5c. One eyewitness is shown a group of people including the defendant and

identifies the defendant as the culprit AND another eyewitness is shown only

one person, the defendant, and the eyewitness identifies the defendant as the

culprit.

| | **50** | 18 | 5 | 4 | 18 | 2 | 3 |

5d. One eyewitness is of the same race as the person being identified as the

culprit, AND another eyewitness is of a different race from the person being

identified as the culprit.

| | **21** | **15** | 4 | 3 | 48 | 5 | 4 |

6a. Eyewitnesses often are asked to estimate how much time elapsed during the commission of the crime. Do

46 JURIMJ 177
46 Jurimetrics J. 177

you think eyewitness estimations of the duration of crimes are usually accurate or usually NOT accurate?

Usually accurate	29	**Skip to Q.7**
Usually not accurate	**58**	**CONTINUE**
Not sure	13	**Skip to Q.7**

(Ask only of respondents who say eyewitness estimates of crime duration are usually not accurate in Question 6a.)
6b. What do you think happens more often--that an eyewitness UNDERESTIMATES the time that elapsed during the commission of the crime and says that the event took LESS time than it actually did, or that an eyewitness OVERESTIMATES the time that elapsed during the commission of the crime and says that the event took MORE time than it actually did?

More likely that an eyewitness underestimates the time	14
More likely that an eyewitness overestimates the time	37
Not sure	7
Respondents who say eyewitness estimates of crime duration are usually accurate/not sure (Q.6a)	42

*209 7. In situations in which **(READ ITEM),** do you think this makes an eyewitness' memory about the details of the crime more reliable, less reliable, or has no effect on the eyewitness' memory about the details of the crime?

	More Reliable	Less Reliable	No Effect On Reliability	Not Sure
A weapon is involved in a crime	37	**30**	25	8
A crime is violent	39	**30**	22	9

8. Sometimes police talk to an eyewitness and generate a sketch of the culprit based on that eyewitness' description. Do you think such a sketch is a reliable tool in identifying the culprit or not?

Yes, sketch is a reliable tool	65
No, sketch is not a reliable tool	**25**
Not sure	10

9. Do you think an identification made by an eyewitness who participated in the creation of such a sketch makes that eyewitness' identification more reliable, less reliable, or just as reliable as an identification provided by an eyewitness who did NOT participate in the creation of such a sketch?

More reliable	34
Less reliable	**12**
Just as reliable	48
Not sure	6

*210 Ask Questions 10a and 10b to one-half of the respondents respectively.
10a. There are different ways of conducting identification procedures. One type of identification procedure involves showing an eyewitness a lineup, which may involve a number of individuals standing in a line facing the eyewitness. For each of the following pairs of lineup procedures that I describe to you, procedure **A** and procedure **B**, please tell me which one you think is more reliable, or whether you think both are equally reliable.

	A is more reliable	**B** is more reliable	**A & B** are equally reliable	Not Sure
Procedure A: A lineup where the eyewitness is instructed that the criminal				
suspect may or may not be included in the lineup, OR				

Procedure B: A lineup where the eyewitness is NOT instructed that the criminal

suspect may or may not be included in the lineup.

| 28 | 51 | 15 | 6 |

Procedure A: A witness views a lineup of potential suspects standing next to

one another, OR

Procedure B: A witness views potential suspects one at a time.

| 53 | 24 | 16 | 7 |

Procedure A: A lineup where the police officer running the lineup is AWARE

of who the suspect is, OR

Procedure B: A lineup where the police officer running the lineup is

UNAWARE of who the suspect is.

| 18 | 55 | 21 | 6 |

10b. There are different ways of conducting identification procedures. One type of identification procedure involves showing the eyewitness photographs, which is usually called a photo array. For each of the following pairs of photo array procedures that I describe to you, procedure **A** and procedure **B**, please tell me which one you think is more reliable, or whether you think both are equally reliable.

| **A** is more reliable | **B** is more reliable | **A & B** are equally | Not Sure |

reliable

Procedure A: A photo array where the eyewitness is instructed that the criminal

---

suspect may or may not be included in the lineup, OR

Procedure B: A photo array where the eyewitness is NOT instructed that the

---

criminal suspect may or may not be included in the lineup.

.

| **30** | 51 | 16 | 3 |

Procedure A: A witness views a photo array of potential suspects all at the same

---

time, OR

Procedure B: A witness views a photo array of potential suspects one picture at a

---

time.

| 38 | **39** | 19 | 4 |

Procedure A: A photo array where the police officer running the photo array is

---

AWARE of who the suspect is, OR

Procedure B: A photo array where the police officer running the photo array is

---

UNAWARE of who the suspect is.

| 30 | **48** | 18 | 4 |

*211 11. Now I am going to read you a series of statements and I'd like you to tell me whether you believe each one is true or false.

	T	F	Not Sure
11a. An eyewitness' level of confidence in his or her			
identification is an excellent indicator of that	39	**55**	6
eyewitness' reliability.			
11b. The act of remembering a traumatic event is like a video			
recording in that one can recall details as if they had been	46	**48**	6
imprinted or burned into one's brain.			
11c. Once an eyewitness learns from police that the person			
they identified as the culprit is the suspect the police believe			
committed the crime, that eyewitness is more likely to	**85**	11	4

express greater confidence in their identification than they			
did beforehand.			
11d. An eyewitness who identifies the same culprit in a	89	9	2
number of identification procedures can still be mistaken.			
11e. If more than two eyewitnesses identified the defendant			
as the culprit, it always means the eyewitnesses picked the	19	79	2
right person.			
11f. Eyewitnesses can believe they remember details about a			
crime that they actually learned about later from someone	80	16	4
else, such as the police.			
11g. Eyewitnesses will sometimes identify a person as the			

culprit because they have seen that person somewhere before	**73**	21	6
and the face is familiar, even though the person was not who			
they actually saw committing the crime.			
11h. Before an eyewitness identifies a defendant at trial, if			
they learn that someone else has also identified the defendant			
as the culprit, that eyewitness is more likely to express	**86**	10	4
greater confidence in their identification when they testify in			
front of the jury.			
11i. Generally, eyewitnesses are equally accurate when			
identifying someone of a different race as when they are	27	**66**	7

identifying someone of
their own race.

11j. If an eyewitness was under high stress at the time of the			
crime, the eyewitness will have better recall for the details of	14	**80**	6
the event.			

*212 **Factual Questions:**
F1. What is your age? I am going to read you some age categories. Stop me when we get to your category.

18 - 24	11
25 - 29	9
30 - 34	13
35 - 39	9
40 - 44	12
45 - 49	6
50 - 54	8
55 - 59	8
60 - 64	7

65 - 69	6
70 - 74	5
75 and over	4
Refused	2

F2. Are you currently employed? **(If "currently employed," ask:)** What type of work do you do? **(If "not currently employed," ask:)** Are you a student, a homemaker, retired, or unemployed and looking for work?

Currently Employed

Professional, manager	29
White-collar worker	25
Blue-collar worker	10
Farmer, rancher	-

Not Currently Employed

Student	5
Homemaker	3
Retired	20

Unemployed, looking for work	6
Other	-
Not sure	2

*213 F5. Generally speaking, do you think of yourself as a Republican, a Democrat, an independent, or something else? **(If "Republican" or "Democrat," ask:)** Do you consider yourself a strong (Republican/Democrat) or a not so strong (Republican/Democrat)? **(If "Independent," ask:)** Would you say that you lean more toward the Republicans or more toward the Democrats?"

Strong Republican	4
Not so strong Republican	3
Independent, leans Republican	2
Independent	15
Independent, leans Democrat	7
Not so strong Democrat	12
Strong Democrat	47
Something else/Other	7
Not sure	3

F6. Have you ever been a juror in a CRIMINAL trial? **(If "Yes," ask:)** Was that trial in Washington, D.C.?

Yes, have been a criminal juror in Washington, D.C.	37
Yes, have been a criminal juror outside Washington, D.C.	4
No, have not been a criminal juror	58
Not sure/refused	1

F7. Are you from a Hispanic or Spanish-speaking background?

Yes, Hispanic	5
No, not Hispanic	93
Not sure/refused	2

F8. What is your race--white, black, Asian, or something else?

White	32
Black	61
Asian	2
Other	1
Hispanic (VOL)	2
Not sure/refused	2

---

*214 F9. Last year, what was your total family income from all sources, before taxes? Just stop me when I get to the right category.

Less than $20,000	12
Between $20,000 and $30,000	14
Between $30,000 and $40,000	11
Between $40,000 and $50,000	9
Between $50,000 and $75,000	13
Between $75,000 and $100,000	8
More than $100,000	13
Not sure/refused	20

F10. Have you or a close family member ever been the victim of a crime? **(If "Yes," ask:)** Was the crime a violent or a nonviolent crime?

Yes, have been the victim of a violent crime	27
Yes, have been the victim of a nonviolent crime	25
Yes, have been the victim of both a violent and a nonviolent crime (VOL)	6
No, have not been the victim of a crime	40

46 JURIMJ 177
46 Jurimetrics J. 177

Not sure/refused	2

F11. Have you or a close personal friend or a family member ever been arrested? **(If "Yes," ask:)** And was that you, your friend, or your family member?

Yes, respondent has been arrested	12
Yes, close personal friend has been arrested	12
Yes, family member has been arrested	23
No, have not been arrested	56
Not sure/refused	5

[FNa1]. Richard S. **Schmechel** is Soros Advocacy Fellow with the Public Defender Service for the District of Columbia. Timothy P. O'Toole is Chief, Special Litigation Division with the Public Defender Service for the District of Columbia. Catharine Easterly is Staff Attorney, Special Litigation Division with the Public Defender Service for the District of Columbia. Elizabeth F. Loftus is Distinguished Professor at the University of California, Irvine.

[FN1]. Press Release, Nat'l Sci. Found., False Identification: New Research Seeks to Inoculate Eyewitnesses Against Errors (Jan. 3, 1997), available at http://www.nsf.gov/pubs/stis1997/pr971/pr971.txt.

[FN2]. Several excellent resources are available that introduce this research to nonexperts in more detail. *See* BRIAN L. CUTLER, EYEWITNESS TESTIMONY: CHALLENGING YOUR OPPONENT'S WITNESS (2002); BRIAN L. CUTLER & STEVEN D. PENROD, MISTAKEN IDENTIFICATION: THE EYEWITNESS, PSYCHOLOGY AND THE LAW (1995); ELIZABETH F. LOFTUS, EYEWITNESS TESTIMONYYYY (1996). The American Psychology-Law Society has assembled a comprehensive research bibliography at http://www.ap-ls.org/links/publishingEyewitness.html.

[FN3]. Charles A. Morgan, III et al., *Accuracy of Eyewitness Memory for Persons Encountered During Exposure to Highly Intense Stress*, 27 INT'L J.L. & PSYCHIATRY 265, 265-67 (2004).

[FN4]. *Id.* at 265.

[FN5]. *Id.*

[FN6]. *Id.* at 269-70.

46 JURIMJ 177
46 Jurimetrics J. 177

[FN7]. *Id.*

[FN8]. *See* Kenneth A. Deffenbacher et al., *A Meta-Analytic Review of the Effects of High Stress on Eyewitness Memory*, 28 LAW & HUM. BEHAV. 687, 699 (2004).

[FN9]. Saul Kassin et al., *On the "General Acceptance" of Eyewitness Testimony Research: A New Survey of the Experts*, 56 AM. PSYCHOLOGIST 405, 413-14 (2001).

[FN10]. CUTLER & PENROD, *supra* note 2, at 68.

[FN11]. *Id.* at 19.

[FN12]. The experience of Jennifer Thompson, a twenty-two-year-old college student who was raped in 1984 in her own home, provides one example. *See* Jennifer Thompson, Op-Ed., *I Was Certain, But I Was Wrong*, N.Y. TIMES, June 18, 2000, § 4, at 15. She spoke at length with her assailant. She took every opportunity to study his face as they moved by several locations in the house where lights were on. She studied the man's face, his hairline, his body. A few days after the attack she positively identified a suspect from a photo array. Investigators arrested the individual, and she again picked him out of a live lineup. At both identifications and at trial she stated with complete certainty that the suspect, named Ronald Cotton, was the man who raped her. She was wrong. Eleven years after being incarcerated, Ronald Cotton was exonerated by DNA evidence that also inculpated a man with similar features named Bobby Poole. Poole pled guilty to the crime in 1995. Since that time, Jennifer Thompson has become an outspoken voice for scrutinizing eyewitness identifications with all possible care. Still, despite the evidence and her acceptance of the fact that he was not the one who raped her, Jennifer Thompson's false memories are so real that when she pauses and thinks of her attacker she says she still sees Ronald Cotton. *Frontline: What Jennifer Saw* (PBS television broadcast Feb. 25, 1997). (transcript available at http:// www.pbs.org/wgbh/pages/frontline/shows/dna/interviews/thompson.html).

[FN13]. NATIONAL INSTITUTUTE OF JUSTICE, EYEWITNESS EVIDENCE: A GUIDE FOR LAW ENFORCEMENT (1999).

[FN14]. NATIONAL INSTITUTE OF JUSTICE, EYEWITNESS EVIDENCE: A TRAINER'S MANUAL FOR LAW ENFORCEMENT (2003).

[FN15]. Scott Ehlers, *Eyewitness Identification: State Law Reform*, THE CHAMPION, Apr. 2005, at 34.

[FN16]. The extent of reform being considered in these jurisdictions varies widely. Some follow recommendations from scientists laid out in the American Psychology-Law Society (APLS) recommendations (a division of the American Psychological Association), others use modifications of the APLS recommendations by legal organizations such as the American Bar Association or Department of Justice guidelines.

[FN17]. Today, all but two states have adopted rules of evidence identical or substantially similar to FRE 702. For a survey of state court rules of scientific admissibility through December 15, 1997, *see* Heather G. Hamilton, Note, *The Movement from Frye to Daubert: Where Do the States Stand?*, 38 JURIMETRICS J. 201, 210-13 (1998). Since Ms. Hamilton's note was published, Connecticut and Pennsylvania have adopted comparable language. *See* CONN. CODE EVID. § 7-2 and PA. R. EVID. 702. Only New York and Massachusetts have substantially different rules.

[FN18]. FED. R. EVID. 702. The Advisory Committee characterizes the ability to "assist the trier of fact" as a question whether the "untrained layman would be qualified to determine intelligently and to the best possible degree the particular issue ...." FED. R. EVID. 702 advisory committee's note. Prior to amendment in 2000, the rule stated: "If scientific, technical, or other specialized knowledge will assist the trier of fact to understand the evidence or to determine a fact in issue, a witness qualified as an expert by knowledge, skill, experience, training, or education, may testify thereto in the form of an opinion or otherwise." FED. R. EVID. 702 (1999). The amendment was not intended to make any substantive change to the rule as interpreted by the federal

46 JURIMJ 177
46 Jurimetrics J. 177

courts. *See* FED. R. EVID. 702 advisory committee's note.

[FN19]. *See, e.g.,* FED. R. EVID. 702 advisory committee's note on proposed revision stating:
> This revision is intended to limit the use, but increase the utility and reliability, of party-initiated opinion testimony bearing on scientific and technical issues. The use of such testimony has greatly increased since enactment of the Federal Rules of Evidence. This result was intended by the drafters of the rule, who were responding to concerns that the restraints previously imposed on expert testimony were artificial and an impediment to the illumination of technical issues in dispute.

Id. (citations omitted).
[FN20]. 410 U.S. 284 (1973).

[FN21]. *Id.* at 302.

[FN22]. 470 U.S. 68 (1985).

[FN23]. 476 U.S. 683 (1986).

[FN24]. *Ake,* 470 U.S. at 82.

[FN25]. *Crane,* 476 U.S. at 691.

[FN26]. 530 U.S. 466 (2000).

[FN27]. Bloodsworth v. State, 512 A.2d 1056 (Md. 1986).

[FN28]. Justice for All Act of 2004, 42 U.S.C. § 14136(e) (2004).

[FN29]. *Bloodsworth,* 512 A.2d at 1063:
> I am concerned that the possibility of admitting the evidence would tend to confuse or mislead the jury. This is not just a matter of usurping the province of the jury, although it is in my judgment most certainly that, it is also that such testimony is of little value in aiding the jury in this case. I'm not persuaded that the testimony will be helpful to the jury in understanding the evidence in this case.

[FN30]. Shockingly, the case affirming Mr. Bloodsworth's conviction remains controlling law on the admissibility of eye-witness expert testimony in Maryland and is an egregious example of the widespread judicial resistance to this particular sort of expert testimony. Further details about Mr. Bloodsworth's wrongful conviction may be found in TIM JUNKIN, BLOODSWORTH: THE TRUE STORY OF THE FIRST DEATH ROW INMATE EXONERATED BY DNA (2004).

[FN31]. Sacco v. Massachusetts, 275 U.S. 574 (1927).

[FN32]. 388 U.S. 218 (1967).

[FN33]. Justice Brennan, writing for the majority, quoted Justice Frankfurter:
> The vagaries of eye-witness identification are well known: the annals of criminal law are rife with instances of mistaken identification. Mr. Justice Frankfurter once said: "What is the worth of identification testimony even when uncontradicted? The identification of strangers is proverbially untrustworthy. The hazards of such testimony are established by a formidable number of instances in the records of English and American trials.

46 JURIMJ 177
46 Jurimetrics J. 177

Id. at 228 (citations omitted).

[FN34]. Richard A. Wise & Martin A. Safer, *A Survey of Judges' Knowledge and Beliefs About Eyewitness Testimony*, 40 CT. REV. 6 (2003).

[FN35]. *Id.* at 8.

[FN36]. Occasionally there are skirmishes to get jury instructions about factors relevant to eyewitness reliability; however, very few of these decisions have received appellate review. *But see* State v. Cromedy, 727 A.2d 457, 458-49 (N.J. 1999) (dealing with cross-racial identification jury instruction).

[FN37]. As of fall 2005, approximately ten appellate courts, in addition to the United States Supreme Court, had not given separate consideration to eyewitness identification experts and relied on rulings concerning expert witnesses more generally. There have been no rulings by the highest appellate courts in the following states: Hawaii, Michigan; Mississippi; New Hampshire; New Mexico; North Carolina (but one recent appellate case used a pure discretionary standard, State v. Cole, 556 S.E.2d 666, 670 (N.C. Ct. App. 2001)); Oklahoma (but eyewitness experts have been admitted by a lower appellate court, Bristol v. State, 764 P.2d 887, 890 (Okla. Crim. App. 1988)); Oregon (but admission appears to be discretionary, State v. Fox, 779 P.2d 197, 197-98 (Or. Ct. App. 1989)); Virginia (but appears to be admitted in narrow circumstances, Currie v. Commonwealth, 515 S.E.2d 335, 338-39 (Va. Ct. App. 1999)); Wisconsin (but appears not to be admitted except in narrow circumstances, State v. Sullivan, No. 02-2775-CR, 2003 WL 22091937, at *3 (Wis. Ct. App. Sept. 3, 2003).

[FN38]. As of fall 2005, approximately five federal circuits and twenty-three states and the District of Columbia held that the admissibility of eyewitness identification experts depends on the exercise of trial court discretion without giving any general rules as to what facts might constitute an abuse of discretion: United States v. Stokes, 388 F.3d 21 (1st Cir. 2004), *rev'd on other grounds*, 125 S. Ct. 1678 (2005); United States v. Welch, 368 F.3d 970 (7th Cir. 2004), *rev'd on other grounds*, 125 S. Ct. 1063 (2005); Hager v. United States, 856 A.2d 1143 1147-48 (D.C. 2004); United States v. Smith, 156 F.3d 1046, 1053 (10th Cir. 1998); United States v. Jackson, 50 F.3d 1335, 1340 (5th Cir. 1995); United States v. Rincon, 28 F.3d 921, 923 (9th Cir. 1994); Parker v. State, 968 S.W.2d 592, 596-97 (Ark. 1998) (further narrowing Utley v. State, 826 S.W.2d 268 (Ark. 1992)); Campbell v. People, 814 P.2d 1, 7 (Colo. 1991) (en banc); State v. McClendon, 730 A.2d 1107, 1114 (Conn. 1999); Garden v. State, 815 A.2d 327 (Del. 2003), *rev'd on other grounds*, Garden v. State, 844 A.2d 311 (Del. 2004); McMullen v. State, 714 So. 2d 368, 372 (Fla. 1998); Johnson v. State, 526 S.E.2d 549, 552-53 (Ga. 2000); State v. Alger, 764 P.2d 119, 127-28 (Idaho Ct. App. 1988) (distinguishing the only state supreme court ruling on the subject that occurred prior to a change in evidence rules, State v. Hoisington, 657 P.2d 17 (Idaho 1983)); People v. Enis, 743 N.E.2d 1 (Ill. 2000); State v. Schutz, 579 N.W.2d 317 (Iowa 1998); State v. Kelly, 752 A.2d 188, 191 (Me. 2000); Bloodsworth v. State, 512 A.2d 1056, 1064 (Md. 1986); Commonwealth v. Zimmerman, 804 N.E.2d 336, 340 (Mass. 2004); State v. Miles, 585 N.W.2d 368, 371 (Minn. 1998); State v. Long, 575 A.2d 435, 463 (N.J. 1990); People v. Lee, 750 N.E.2d 63, 65 (N.Y. 2001);State v. Fontaine, 382 N.W.2d 374, 377 (N.D. 1986); State v. Werner, 851 A.2d 1093, 1100 (R.I. 2004); State v. McCord, 505 N.W.2d 388, 391 (S.D. 1993); Weatherred v. State, 15 S.W.3d 540, 543 (Tex. Crim. App. 2000); State v. Percy, 595 A.2d 248, 252-53 (Vt. 1990); State v. Cheatam, 81 P.3d 830, 840 (Wash. 2003); State v. Taylor, 490 S.E.2d 748, 754 (W. Va. 1997); Engberg v. Meyer, 820 P.2d 70, 80 (Wy. 1991).

[FN39]. United States v. Smith, 156 F.3d 1046, 1053 (10th Cir. 1998); Campbell v. People, 814 P.2d 1, 8-9 (Colo. 1991) (en banc); Johnson v. State, 526 S.E.2d 549, 552-53 (Ga. 2000); State v. Cheatam, 81 P.3d 830, 842 (Wash. 2003).

[FN40]. *See, e.g.,* McMullen v. State, 714 So. 2d 368, 371-72 (Fla. 1998) (upholding an earlier decision by that court that espoused a discretionary approach while saying a jury is never in need of an eyewitness expert's advice).

[FN41]. One federal circuit and ten states have articulated some kind of fixed rules regarding what factual circumstances require the admission of eyewitness experts: United States v. Mathis, 264 F.3d 321, 340 (3d Cir. 2001); Skamarocius v. State, 731 P.2d 63, 67 (Alaska Ct. App. 1987); State v. McCutcheon, 781 P.2d 31, 34 (Ariz. 1989) (narrowing State v. Chapple, 660 P.2d 1208 (Ariz. 1983)); People v. Jones, 70 P.3d 359, 374 (Cal. 2003) (following People v. McDonald, 690 P.2d 709 (Cal. 1984)); Cook v. State, 734 N.E.2d 563, 570 (Ind. 2000); Commonwealth v. Christie, 98 S.W.3d 485, 490 (Ky. 2002); State v. DuBray,

77 P.3d 247, 255 (Mont. 2003); White v. State, 926 P.2d 291, 293 (Nev. 1996); State v. Buell, 489 N.E.2d 795, 803 (Ohio 1986); State v. Frazier, 592 S.E.2d 621, 623 (S.C. 2004) (following State v. Whaley, 406 S.E.2d 369 (S.C. 1991)); State v. Maestas, 63 P.3d 621, 626 (Utah 2002).

[FN42]. State v. Chapple, 660 P.2d 1208, 1223-24 (Ariz. 1983); People v. McDonald, 690 P.2d 709, 723 (Cal. 1984), *superceded by* 738 P.2d 764 (1987); United States v. Downing, 753 F.2d 1224, 1241 (3d Cir. 1985).

[FN43]. State v. McCutcheon, 781 P.2d 31, 34 (Ariz. 1989).

[FN44]. White v. State, 926 P.2d 291 (Nev. 1996).

[FN45]. Three federal circuits and three states have indicated that eyewitness experts should be excluded except in special circumstances: United States v. Martin, 391 F.3d 949, 954 (8th Cir. 2004) (following United States v. Kime, 99 F.3d 870 (8th Cir. 1996)); United States v. Bellamy, 26 Fed. Appx. 250, 259 (4th Cir. 2002) (following United States v. Harris, 995 F.2d 532 (6th Cir. 1993)); United States v. Lumpkin, 192 F.3d 280, 289 (2d Cir. 1999); *Ex parte* Williams, 594 So. 2d 1225, 1227 (Ala. 1992); State v. Chapman, 436 So. 2d 451, 453 (La. 1983); State v. Whitmill, 780 S.W.2d 45, 47 (Mo. 1989).

[FN46]. *See* cited cases, *supra* note 38.

[FN47]. United States v. Smith, 122 F.3d 1355, 1357-59 (11th Cir. 1997); State v. Gaines, 926 P.2d 641, 649 (Kan. 1996); State v. George, 645 N.W.2d 777, 790 (Neb. 2002); Commonwealth v. Simmons, 662 A.2d 621, 631 (Pa. 1995); State v. Coley, 32 S.W.3d 831, 833-34 (Tenn. 2000).

[FN48]. Discretionary standards do not inherently work this way. But the combination of two factors lead to this result: (1) lack of guidance about how to exercise discretion; and (2) no finding of abuse in cases excluding experts.

[FN49]. *See, e.g.*, People v. Lee, 750 N.E.2d 63, 66-67 (N.Y. 2001).

[FN50]. *See* Kyles v. Whitley, 514 U.S. 419, 446 (1995) (noting the clear importance of eyewitness testimony even in cases where substantial amounts of other evidence exists: "[t]he effective impeachment of one eyewitness can call for a new trial even though the attack does not extend to others, as we have said before.").

[FN51]. 541 U.S. 36 (2004).

[FN52]. *Id.* at 62.

[FN53]. In the case of exoneree Ronald Cotton, the government introduced a pair of black shoes and a flashlight found at his home into evidence at trial to corroborate Jennifer Thompson's eyewitness identification and description of a man wearing black shoes and possessing a flashlight. Apart from that eyewitness identification, Ronald Cotton's possession of such ordinary items would never have been deemed identification evidence.

[FN54]. Kirk Bloodsworth's case exemplifies how the corroboration of several eyewitnesses that seemed to mutually reinforce each other, in retrospect, may not be trustworthy when the identifications were not truly independent of one another. The same flawed interrogation procedures, investigator bias (or fraud), or postevent information supplied to eyewitnesses may create a false appearance of independent corroboration.

[FN55]. Manson v. Brathwaite, 432 U.S. 98, 118 & n.* (1977) (Stevens, J., concurring) (warning against reliance on corroborating evidence when making decisions about the admissibility of identification testimony).

46 JURIMJ 177
46 Jurimetrics J. 177

[FN56]. FED. R. EVID. 702.

[FN57]. Wise & Safer, *supra* note 34, at 8.

[FN58]. *Id.*

[FN59]. *See, e.g.,* Hager v. United States, 856 A.2d 1143, 1148-49 (D.C. 2004); Skamarocius v. State, 731 P.2d 63, 66 (Alaska Ct. App. 1987); Commonwealth v. Christie, 98 S.W.3d 485, 490 (Ky. 2002).

[FN60]. *Compare* Garden v. State, 815 A.2d 327 (Del. 2003) (*rev'd on other grounds*, Garden v. State, 844 A.2d 311 (Del. 2004)) (holding that expert testimony on the correlation of confidence and accuracy should be permitted to the same extent as other admissible topics such as cross-racial bias) *and* Skamarocius, 731 P.2d at 66 (finding that confidence is commonly, but mistakenly, thought to correlate with accuracy) *and* United States v. Downing, 753 F.2d 1224, 1230 n.6 (3d Cir. 1985) (noting that confidence-accuracy studies are often counterintuitive to jurors) *with* United States v. Mathis, 264 F.3d 321, 340 (3d Cir. 2001) (holding that the correlation between confidence and accuracy in identifications was not a proper subject for expert testimony).

[FN61]. *Compare* Commonwealth v. Christie, 98 S.W.3d 485, 490 (Ky. 2002) (finding stress to be among the narrow circumstances sufficient to support eyewitness expert admission) *and* United States v. Mathis, 264 F.3d 321, 340 (3d Cir. 2001) (finding stress is among the proper subjects for expert testimony) *with* Green v. United States, 718 A.2d 1042, 1053 (D.C. 1998) (upholding trial finding that effects of stress on eyewitnesses are not beyond the ken of jurors).

[FN62]. *See infra* note 69.

[FN63]. Wise & Safer, *supra* note 34, at 10.

[FN64]. At worst, such opinions are being disingenuous about other motives, such as keeping out evidence that almost always benefits defendants.

[FN65]. This inquiry was an outgrowth of PDS's duty to give the best representation possible to indigent clients and was conducted in the context of several ongoing cases involving purported eyewitness identifications.

[FN66]. A few studies have assessed attorneys' understanding of eyewitness reliability variables. All of these studies, however, involve small samples and some date from the early 1980s. For an overview, see Jennifer Devenport et al., *Eyewitness Identification Evidence*, 3 PSYCHOL. PUB. POL'Y & L. 338 (1997).

[FN67]. *Id.* at 343.

[FN68]. *Compare* Wise & Safer, *supra* note 34, *with* CUTLER & PENROD, *supra* note 2, at 159-68.

[FN69]. Kenneth A. Deffenbacher & Elizabeth F. Loftus, *Do Jurors Share a Common Understanding Concerning Eyewitness Behavior?*, 6 LAW & HUM. BEHAV. 15 (1982); Kevin M. McConkey & Suzanne M. Roche, *Knowledge of Eyewitness Memory*, 24 AUSTL. PSYCHOL. 377 (1989); Elizabeth Noon & Clive R. Hollin, *Lay Knowledge of Eyewitness Behaviour: A British Survey*, 1 APPLIED COGNITIVE PSYCHOL. 143 (1987).

[FN70]. CUTLER & PENROD, *supra* note 2, at 101, 175 Table 11.1 (1995) (noting that even those factors most often recognized by test takers, like cross-racial bias, were missed by 25% or more of the participants).

[FN71]. *Id.* at 346-49.

46 JURIMJ 177
46 Jurimetrics J. 177

[FN72]. Admittedly, surveys are subject to the critique that they cannot assess how jurors are actually able to apply their knowledge about eyewitness identifications in the courtroom--a question to which postdiction and judgment studies may be better suited.

[FN73]. The project was initiated by Edward J. Ungvarsky, Public Defender Service for the District of Columbia (PDS) Special Counsel and a Vice Chair of the NACDL Forensic Evidence Committee, and PDS's Special Litigation Division, a unit that works exclusively on systemic criminal justice issues such as forensic evidence challenges and issues surrounding the suppression of exculpatory evidence by prosecutors. The motivation behind this project was simple: If jurors understand as a matter of common sense what makes some eyewitness identifications more reliable than others, it would not make sense for PDS to continue to devote resources toward educating already-informed jurors on this topic. On the other hand, if jurors actually have an incomplete understanding of the factors that affect eyewitness reliability, it would remain imperative to continue to seek to provide jurors with the tools necessary to ensure that they could intelligently evaluate the evidence presented to them by the government.

[FN74]. *See infra* Appendix.

[FN75]. EDWARD ARNOLDS ET AL., EYEWITNESS TESTIMONY: STRATEGIES AND TACTICS 14-15 (1984); David M. Shofi, The *New York Courts' Lack of Direction and Discretion Regarding the Admissibility of Expert Identification Testimony,* 13 PACE L. REV. 1101, 1104 (1994); Gary L. Wells, *Eyewitness Behavior,* 4 LAW & HUM. BEHAV. 237, 238 (1980).

[FN76]. KARL HABERLANDT, HUMAN MEMORY: EXPLORATION AND APPLICATION 4 (1999) (defining and describing reconstructive memory).

[FN77]. *See, e.g.,* ELIZABETH F. LOFTUS & JAMES M. DOYLE, EYEWITNESS TESTIMONY: CIVIL AND CRIMINAL, § 2.10 (2d ed. 1992); Elizabeth F. Loftus et al., *Some Facts About "Weapon Focus,"* 11 LAW & HUM. BEHAV. 55 (1987); Anne Maass & Gunther Kohnken, *Eyewitness Identification: Simulating the "Weapon Effect,"* 13 LAW & HUM. BEHAV. 397 (1989).

[FN78]. An archival study of actual police cases demonstrated that the presence of a weapon "did not reduce the quantity of descriptive information the witness was subsequently able to provide to police about the culprit, but it did impair witnesses' subsequent ability to recognize the culprit." *See* Donald P. Judges, *Two Cheers for the Department of Justice's Eyewitness Evidence: A Guide for Law Enforcement,* 53 ARK. L. REV. 231, 243 n.37 (2000) (citing Patricia A. Tollestrup et al., *Actual Victims and Witnesses to Robbery and Fraud: An Archival Analysis, in* ADULT EYEWITNESS TESTIMONY: CURRENT TRENDS AND DEVELOPMENTS 144, 158 (David F. Ross et al. eds., 1994)); *see also* Nancy Mehrkens Steblay, *A Meta-Analytic Review of the Weapon Focus Effect,* 16 LAW & HUM. BEHAV. 413 (1992). Experiments in which eye movements are monitored while subjects witness a scene where a weapon is involved also demonstrate the existence of the phenomenon of "weapon focus." Brian L. Cutler et al., *The Reliability of Eyewitness Identification,* 11 LAW & HUM. BEHAV. 233, 240, 244 (1987); Elizabeth F. Loftus et al., *Some Facts About "Weapon Focus,"* 11 LAW & HUM. BEHAV. 55, 57-61 (1987).

[FN79]. Taylor v. United States, 451 A.2d 859, 866-67 n.9 (D.C. 1982).

[FN80]. *E.g.,* Morgan, III et al., *supra* note 3. In another study, it was observed that "[p]erceptual abilities actually decrease in a highly stressful situation, and the person under stress is less reliable than he or she would be otherwise. Such a witness becomes less capable of remembering details, less accurate in reading dials, and less certain in detecting signals." Thomas J. Feeney, *Expert Psychological Testimony on Credibility Issues,* 115 MIL. L. REV. 121, 146 (1987). "One theory indicates that moderate levels of stress or arousal increase performance up to a point. Under this theory, known as the Yerkes-Dodson law, perceptual performance follows a U-shaped curve. At very low levels of arousal, the senses are not yet functioning fully.

46 JURIMJ 177
46 Jurimetrics J. 177

Performance peaks at moderate levels of arousal and then declines as the stress increases further." *Id.* at 146 n.163 (citation omitted).

[FN81]. This is unsurprising. Researchers have postulated that lay people share "a common" but mistaken "belief that stress heightens a witness' observation powers and 'burns' an image of the scene into the mind." Feeney, *supra* note 80, at 146.

[FN82]. *See, e.g.,* ELIZABETH F. LOFTUS & JAMES M. DOYLE, EYEWITNESS TESTIMONY: CIVIL AND CRIMINAL 26-27 (1987) (explaining that a witness tends to overestimate the duration of an especially stressful or violent event); JAMES MARSHALL, LAW AND PSYCHOLOGY IN CONFLICT 41-81 (1966) (on average, viewers estimated that a 42-second film in which man rocks a baby carriage and then flees when a woman approaches lasted a minute and a half). Another study reported that in a survey of over 100 experimental psychologists, potential jurors, judges, law students, and lawyers, 95% of the experts agreed that witnesses generally overestimate the duration of crimes, whereas fewer than half of the potential jurors did so. A. Daniel Yarmey & Hazel P. Tressillian Jones, *Is the Psychology of Eyewitness Identification a Matter of Common Sense?, in* EVALUATING WITNESS EVIDENCE: RESEARCH AND NEW PERSPECTIVES 33 (Sally M. Lloyd-Bostock & Brian R. Clifford eds., 1983). *See also* Elizabeth F. Loftus et al., *Time Went by So Slowly: Overestimation of Event Duration by Males and Females,* 1 APPL. COGNITIVE PSYCHOL. 1 (1987).

[FN83]. In some instances, studies have shown no meaningful correlation between confidence and accuracy. *See, e.g.,* Cutler et al., *supra* note 78, at 233, 234; Kenneth A. Deffenbacher, *Eyewitness Accuracy and Confidence,* 4 LAW & HUM. BEHAV. 243, 258 (1980).

[FN84]. *E.g.,* Amy L. Bradfield et al., *The Damaging Effect of Confirming Feedback on the Relation Between Eyewitness Certainty and Identification Accuracy,* 87 J. APPLIED PSYCHOL. 112 (2002).

[FN85]. Even if survey respondents who said that an identification of a confident witness is "equally reliable" as one by a less confident witness are treated as having correctly understood this phenomenon, 59% of the respondents--more than half of those on any given jury--demonstrate a fundamental misunderstanding of the confidence accuracy relationship.

[FN86]. Elizabeth F. Loftus, *Eyewitnesses: Essential But Unreliable,* 18 PSYCHOL. TODAY 22, 24 (1984).

[FN87]. Brian Clifford, *Police as Eyewitnesses,* 36 NEW SOC'Y 176 (1976) (demonstrating that police officers perform more poorly than civilians as witnesses because police officers often have a biased interpretation of events).

[FN88]. *See generally* ELIZABETH LOFTUS, EYEWITNESS TESTIMONY § 4.9 (3d ed. 1997 & Supp. 1999); GARY L. WELLS & ELIZABETH LOFTUS, PSYCHOLOGICAL PERSPECTIVES 1 (1984); Sheri Lynn Johnson, *Cross-Racial Identification Errors in Criminal Cases,* 69 CORNELL L. REV. 934 (1984); Stephanie J. Platz & Harmon M. Hosch, *Cross-Racial/Ethnic Eyewitness Identification: A Field Study,* 18 J. APPLIED SOC. PSYCHOL. 972 (1988); John P. Rutledge, *They All Look Alike: The Inaccuracy of Cross-Racial Identifications,* 28 AM. J. CRIM. L. 207 (2001).

[FN89]. *See* Stovall v. Denno, 388 U.S. 293, 302 (1967) ("the practice of showing suspects singly to persons for the purpose of identification, not as part of a lineup, has been widely condemned."); UNITED STATES DEPARTMENT OF JUSTICE, EYEWITNESS EVIDENCE: A TRAINER'S MANUAL FOR LAW ENFORCEMENT 30 (2003) (discussing how to reduce the inherent suggestiveness of show-up procedures).

[FN90]. *See* Moore v. Illinois, 434 U.S. 220, 229 (1977) ("Indeed, a one-on-one confrontation generally is thought to present greater risks of mistaken identification than a lineup."); Simmons v. United States, 390 U.S. 377, 383 (1968) ("[The danger of misidentification] will be increased if the police display to the witness only the picture of a single individual who generally resembles the person he saw, or if they show him the pictures of several persons among which the photograph of a single such individual recurs or is in some way emphasized.").

46 JURIMJ 177
46 Jurimetrics J. 177

[FN91]. A. Daniel Yarmey et al., *Accuracy of Eyewitness Identifications in Showups and Lineups*, 20 LAW & HUM. BEHAV. 469, 459-77 (1996).

[FN92]. For studies supporting the relative judgment phenomenon, see Gary L. Wells & Eric P. Seelau, *Eyewitness Identification: Psychological Research and Legal Policy on Lineups*, 1 PSYCHOL. PUB. POL'Y. & L. 765 (1995); Gary L. Wells, *What Do We Know about Eyewitness Identification?*, 48 AM. PSYCHOL. 553 (1993); Gary L. Wells, *The Psychology of Lineup Identifications*, 14 J. APPLIED SOC. PSYCHOL. 89 (1984).

[FN93]. Wells & Seelau, *supra* note 92, at 765, 772-73. *See also* CUTLER & PENROD, *supra* note 2, at 101, 115-23 (concluding that "the research shows that biased instructions substantially increase the likelihood of false identifications"); Roy S. Malpass & Patricia G. Devine, *Eyewitness Identification: Lineup Instructions and the Absence of the Offender*, 66 J. APPLIED PSYCHOL. 482, 482-89 (1981) (failure to give explicit instructions to the eyewitness that explain that the perpetrator might not be in the lineup leads subjects to select someone from the lineup regardless of whether the perpetrator is present; even when these instructions are given, eyewitnesses tend to make relative judgments.).

[FN94]. An equal number (51%) thought an identification from a photo array would be *more* reliable if the eyewitness was *not* instructed about the culprit's potential absence, while an additional 19% either thought it did not matter whether such an instruction was given or were not sure. Similarly, only 30% of the respondents correctly understood that an identification from a photo array is more reliable if the witness is instructed that the actual culprit "may or may not be in the photo array."

[FN95]. *See* R.C.L. Lindsay et al., *Sequential Lineup Presentation: Technique Matters*, 76 J. APPLIED PSYCHOL. 741, 744 (1991); Brian L. Cutler & Steven D. Penrod, *Improving the Reliability of Eyewitness Identification: Lineup Construction and Presentation*, 73 J. APPLIED PSYCHOL. 281, 288-89 (1988); R.C.L. Lindsay & Gary L. Wells, *Improving Eyewitness Identifications from Lineups: Simultaneous Verses Sequential Lineup Presentations*, 70 J. APPLIED PSYCHOL. 556, 562 (1985); *see also* Nancy M. Steblay et al., *Eyewitness Accuracy Rates in Sequential and Simultaneous Lineup Presentations: A Meta-Analytic Comparison*, 25 LAW & HUM. BEHAV. 459, 471 (2001) (meta-analysis of twenty-three papers showed that a sequential lineup, as opposed to a simultaneous lineup, reduced the rate of mistaken identification in culprit-absent lineups but was also associated with a reduction in accurate identification rates in culprit-present lineups).

[FN96]. *See, e.g.*, Ronald P. Fisher, *Interviewing Victims and Witnesses of Crime*, 1 PSYCHOL., PUB. POL'Y, & L. 732, 753-58 (1995) (discussing how interviewers' expectations affect subjects' performance); Gary L. Wells & Amy L. Bradfield, *"Good, You Identified the Suspect": Feedback to Eyewitnesses Distorts Their Reports of the Witnessing Experience*, 83 J. APPLIED PSYCHOL. 360 (1998) (providing studies demonstrating confirming feedback given at the time of an initial identification made subjects significantly more confident of their identification of a suspect).

[FN97]. Gary L. Wells et al., *Eyewitness Identification Procedures: Recommendations for Lineups and Photospreads*, 22 LAW & HUM. BEHAV. 603, 627-29 (1998) (police officers aware of the perpetrator's identity can unconsciously bias an eyewitness' lineup selection).

[FN98]. *See* EYEWITNESS TESTIMONY: PSYCHOLOGICAL PERSPECTIVES (Gary L. Wells & Elizabeth F. Loftus eds. 1984).

46 Jurimetrics J. 177

END OF DOCUMENT

A Meta-Analytic Review of the Weapon Focus Effect

Nancy Mehrkens Steblay

Law and Human Behavior, Vol. 16, No. 4 (Aug., 1992), 413-424.

A Meta-Analytic Review of the Weapon Focus Effect*

Nancy Mehrkens Steblay

This meta-analytic review examined 19 tests of the weapon focus effect—the hypothesis that the presence of a weapon during commission of a crime will negatively affect an eyewitness's ability to later identify the perpetrator. A significant overall difference between weapon-present and weapon-absent conditions was demonstrated, with weapon presence leading to reduced identification accuracy. Overall, the size of the effect was small (.13) for the dependent measure of lineup identification and moderate (.55) for feature accuracy. Discussion focuses on those factors that appear to mediate and facilitate the weapon focus effect.

Research shows that eyewitness testimony can have a strong impact on juries (Loftus, 1974). Offered as direct evidence, positive identification of a perpetrator is uniquely powerful in its evidentiary status (Wells, 1985). Unintentional errors in eyewitness identification can occur, however, with possible miscarriages of justice involving incrimination of the wrong party and/or failure to identify the guilty party. Assessment of the accuracy of eyewitness testimony is a weighty charge for the court, and psychologists with knowledge relevant to eyewitness processes— perception, information processing, and memory—are in a position to help with this task. The usefulness of psychological knowledge in this arena and the appropriateness of psychologists in the role of advisor to the court, however, depend heavily on an adequate base of knowledge of eyewitness-relevant variables. One of these variables, termed *weapon focus*, is the subject of this investigation, and a statement about the base of knowledge regarding this variable is the goal of the project.

Weapon focus refers to the visual attention that eyewitnesses give to a perpetrator's weapon during the course of a crime. It is expected that the weapon will draw central attention, thus decreasing the ability of the eyewitness to adequately encode and later recall peripheral details. Research efforts have assessed eyewitness recall of various crime details in an attempt to establish the parameters of weapon focus effects on perception and memory (e.g., Kuehn, 1974). Most tests of the weapon focus effect have included a weapon's presence or absence during a crime as the independent variable and the subject's later attempt to identify the perpetrator in a lineup as the dependent measure. The experimental hypothesis has been that the frequency of correct identifications will be greater in the weapon-absent condition.

A quick tally of results from these studies indicates that six tests show clear support of the weapon focus effect (at the traditional .05 level of significance), no tests suggest a significant opposite effect (i.e., that weapon presence increases accurate identification) and 13 tests report no significant differences between weapon-present and -absent conditions. Thus, although there is support for the weapon focus effect, the number of nonconforming research results may lead to concern about the reliability of the phenomenon. Indeed, through a recent survey, Kassin, Ellsworth, and Smith (1989) found that only 56.6% of experts in this field reported that the weapon focus effect is reliable enough for psychologists to present in courtroom testimony.

A closer examination of these research results may more fully explain the weapon focus phenomenon. The presence (or absence) of certain factors in supportive studies may provide insight into the causal nature of the phenomenon and provide explanation as to when and why weapon focus may be detrimental to eyewitness performance.

This investigation is a statistical review of all available literature. Twelve studies have been located that address the weapon focus effect; these include 19 tests of the hypothesis. These studies examine conceptually similar variables and allow reliable application of the statistical procedures used in meta-analysis. Among other information, meta-analysis allows computations of the overall probability that the pattern of results in a set of studies was due to chance, an overall effect size estimate, and a figure regarding the number of additional nonsupportive studies needed to reverse the meta-analytic conclusions.

One useful outcome of such statistical conclusions is a quantitative indicator regarding the status of the weapon focus effect. As noted earlier, psychologists who take the role of expert witness need a base of knowledge upon which to draw for their testimony. As a practical and ethical issue, supporting research must be available to validate expert psychological testimony. Legally, the expert testimony must also meet the criteria of the *Frye* test (*Frye v. United States,* 1923, p. 1014.). Following *Frye,* the testimony should conform to a "generally accepted explanatory theory" (*U.S. v. Amaral,* 1973, p. 1153). The advantage of meta-analysis is that researchers generally agree on the meaning and usefulness of probability and effect size indicators. Thus the *Frye* test can be informed with a standard statistical indicator.

The goals of the present review may be summarized as follows: (1) to exam-

ine the weapon focus literature, defining (a) the overall probability of a weapon present–absent difference in identification accuracy, (b) the size of this effect, and (c) a confidence level that the research reviewed accurately represents all studies conducted; (2) to examine a number of factors that relate to and may help to explain the weapon focus effect; and (3) to identify gaps in our knowledge and necessary areas for future research.

METHOD

Sample

Several steps were taken to locate potential studies, including a computer search of two abstract services, PsycINFO and SOCIAL SCISEARCH. A manual search of Psychological Abstracts and the library book catalog was also conducted. Subsequently, a complete search of relevant references cited in any of the previously located articles was performed. Finally, authors of primary articles were contacted by mail and asked to send any relevant works, published or unpublished. The final sample included nine published reports, two convention papers, and one unpublished dissertation (Kramer, personal communication, September, 1990). Possible selection bias in published studies was a concern and, as will be discussed later, was partially addressed through calculation of a fail-safe N. An individual study was selected for inclusion in the review if the author(s) used a weapon present/absent manipulation (including studies with high versus low visibility of weapon) with a dependent measure of lineup identification. The final sample consisted of 19 sets of data testing the weapon focus hypothesis, from 12 empirical reports representing 2,082 subjects. The review covered those studies available as of March, 1991 (see Appendixes A and B).

Study Characteristics

Several methodological and theoretical characteristics were recorded. Among these were source (published, convention paper, unpublished manuscript), author, number of hypothesis tests per study, sample size, variable definitions, specifics of research design, and supplemental dependent measures.

Statistical Analyses

Z-Score

Following the work of Rosenthal (1984), an overall probability level associated with the observed pattern of results was calculated by combining Z-scores of individual tests of the hypothesis. Exact one-tailed probability levels were calculated for each test through reconstruction of the data, by returning to other statistics reported within the article, or by obtaining this information directly from authors of the articles. Recovery of sample sizes and proportion of correct identifications per condition allowed calculation of Z-scores for the difference be-

tween proportions. For one test, retrieval of the data was not possible, and a conservative policy (Rosenthal, 1984) was followed: The report of a significant main effect with no statistics available was constructed as $Z = 1.65$, $p = .05$.

Following the technique of Beaman, Cole, Klentz, Preston, and Steblay (1983), three summary standard normal deviates (Z-scores) were calculated. The overall Z-score (Z_{ma}) provides an unweighted estimate of the overall probability level. In addition, a Z-score was calculated that included weighted individual test Z's (Z_{mn}). This weighting by sample size of the study provides an estimate of population parameters that allows greater emphasis on larger samples (with accordingly more reliable estimates). A final weighted Z (Z_{mf}) was computed using fractional weights equivalent to the reciprocal of the number of tests contained within each study. This weighting technique essentially adjusts for the nonindependence of hypothesis tests within single studies.

Mean Effect Size

The effect size computed for each individual hypothesis test is the coefficient recommended by Cohen (1977) for use when testing differences between proportions. To calculate an overall effect size, individual effect sizes were weighted by the reciprocal of the effect size variance and the summation of these values divided by the sum of the weights. This procedure is considered analogous to that offered by Hedges (1984) for mean differences and is appropriate for effect sizes based on differences between arc sin transformed proportions (see Beaman et al., 1983). This procedure is necessary to adjust for the different sample sizes which affect accuracy of effect size indicators. In the remainder of this article, reference to "mean effect size" denotes the weighted mean computed as above.

Fail-Safe N

In conjunction with the computation of the meta-analytic normal deviate, a fail-safe N was calculated following the procedure of Rosenthal (1978). This method results in an estimate of the number of additional tests averaging null results (Z-score of zero) that would be needed in order to bring the significance level attained through the meta-analysis to a value larger than .05. When appropriate, a "file drawer" analysis based on effect size was also calculated (Hunter & Schmidt, 1990). This figure reports the number of unlocated studies averaging null results needed to reduce the established effect size for a set of studies to a chosen (nonsignificant) value.

RESULTS

Meta-Analysis 1: Identification Accuracy

The 19 experimental tests of the weapon focus hypothesis yield a Z_{ma} of 3.40 ($p = .0003$), indicating greater lineup identification accuracy for the weapon-absent condition than the weapon-present condition. Weighting the individual

WEAPON FOCUS

tests by their sample sizes produce a Z_{mn} of 2.65 ($p = .004$). A weighting based on the number of tests per study generates $Z_{mf} = 2.24$ ($p = .01$). These significant results are further underscored by calculation of the fail-safe N (N_{fs}). For Z_{ma}, this figure is 1,523, which suggests that this large number of unlocated and unsupportive studies would need to exist before the p figure would change to a nonsignificant level. This well exceeds the tolerance level (105) computed as directed by Rosenthal (1984). Even the smallest Z derived, Z_{mf}, still generates a respectable N_{fs} of 280.

The mean effect size for the group of 19 tests is .13 [CI (95%) = .01 to .25]. This is a relatively small effect size; in practical terms this suggests that the area of nonoverlap (Cohen, 1977) between the distributions of the weapon-present and -absent conditions is 9.8%. A file drawer analysis based on effect size indicates that 216 uncovered studies averaging null results are needed to reduce the effect size to .01 (approximating zero).

Meta-Analysis 2: Feature Accuracy

Ten of the weapon focus data sets included dependent measures beyond the lineup identification test. Although the form of the dependent measures varied somewhat from study to study, the existing commonality is a series of questions posed to subjects that assessed each subject's memory of perpetrator characteristics peripheral to the weapon. For example, subjects were asked to describe the target's clothing or facial features.

The 10 studies that reported comparisons between weapon-absent and -present conditions for feature accuracy scores produce a Z_{ma} of 5.88 ($p < .0001$), establishing that the weapon-absent condition generated significantly more accurate descriptions of the perpetrator than did the weapon-present condition. The fail-safe N for this group of studies is 1,263, and the overall effect size, now calculated as d, is .55 [CI (95%) = .24 to .86].[1]

Meta-Analyses of Subsets of Data

It is valuable to go beyond the original primary analyses to seek information regarding relationships among various subsets of the data. This allows investigation into both methodological issues and theoretical underpinnings of the hypothesized weapon focus effect. It is particularly important to account for the range of effect sizes ($-.43$ to .67) in this set of studies. This was attempted through examination of the following variables.

Author of Publication

The lower end of an effect size frequency distribution, specifically including outcomes of zero or negative effect size, is comprised of work by Bothwell,

---

[1] Hedges's (1984) correction applied.

Kramer, and Cutler and colleagues.[2] Each of these author subsets was reviewed separately in an attempt to isolate methodological factors that differentiate these studies and that might shed light on the causal nature of the weapon focus effect.

The work of Bothwell includes at least two characteristics that may relate to a reduced effect size. First, Bothwell used a video presentation as the stimulus for his subjects. Second, and perhaps more importantly, the video scenario did not portray a crime; rather, a single actor simply carried a weapon as he walked down a corridor. The Z_{ma} generated for these two tests of the hypothesis is $-.79$, $h = -.22$.[3]

Five of the six hypothesis tests of Kramer et al. are also represented on the lower end of the effect size continuum ($Z_{ma} = .68$, $p = .25$, $h = .08$). This work similarly includes noteworthy methodological aspects: These five tests were designed to create a very nonarousing, environmentally stark scenario. No crime is actually committed, no other bystanders surround the actor, and the "weapon" (a bloody meat cleaver) is not used as a weapon, but rather just carried by the actor. Also, Kramer et al.'s scenario is presented on slides. An interesting contrast in scenario is provided by a sixth test of the hypothesis by Kramer et al. (actually the first reported in the 1990 article). This was also a slide presentation, but a crime was enacted, complete with sound effects. The Z_{ma} for this single test was 1.64 ($p = .05$, $h = .61$). At this point it should be noted, also, that although five tests from Kramer et al.'s work produce quite low effect sizes for lineup identification, feature accuracy scores show significant differences between conditions, with a combined effect size of .79.

The work of Cutler and colleagues (which also includes the O'Rourke article) provides six tests of the hypothesis and generates a significant Z_{ma} (1.62, $p = .05$) and an effect size of .11, only slightly less than the complete sample. Cutler's work, like Bothwell's involved a video scenario; an important difference is that a crime was enacted in the Cutler video. An additional characteristic may also be important: The weapon (a gun) was not entirely absent in the weapon-absent condition, but rather in the pocket of the perpetrator.

Having identified potential explanatory variables from these subsets, subsequent analyses address these variables and others across the complete set of studies.

Weapon Visibility

The most direct test of the weapon focus hypothesis, as previously stated, would involve the absence versus presence of a weapon. The operational definition of weapon absence varied from one project to another, resulting in three primary subgroups. The absent weapon in the Cutler studies actually involved a weapon present in the perpetrator's pocket; $Z_{ma} = 1.62$, $h = .11$. One test

---

[2] In order to simplify reporting of results, only the first author of research articles included in the meta-analysis will be named from this point on.

[3] Z_{mf} and Z_{mn} were also calculated for these and the following groups. However, such figures are reported only if they differ from Z_{ma} in a noteworthy manner.

WEAPON FOCUS

(Kramer, 1990) involved a weapon visible in hand (a bottle), but not used as a weapon in the weapon-absent condition ($Z_{ma} = 1.64$, $h = .61$). Studies in which the weapon was truly absent from the subject's view ($N = 12$) produced $Z_{ma} = 2.67$, $p = .0038$, $h = .12$.

Mode of Presentation

In presenting the stimulus scenario to subjects, researchers used one of three general strategies. The Cutler and Bothwell research used a video mode ($Z_{ma} = 1.01$, $p = .16$, $h = -.10$). Johnson and Maass used actors to simulate a real-life event ($Z_{ma} = 2.25$, $p = .01$, $h = .38$). The remaining researchers ($N = 9$) used a slide presentation. ($Z_{ma} = 2.93$, $p = .002$, $h = .25$).

Arousal Level

The studies can also be considered as three separate groups based on the potential for emotional arousal in the experimental scenario. Assuming that the "real-life" scenarios of Johnson and Maass can be considered to be potentially more arousing than the remaining tests, these two can be compared to the others. As noted above, $Z_{ma} = 2.25$, $p = .01$, $h = .38$. Studies in which a crime scene with a recognizable weapon is viewed via video or slides ($N = 10$) can be considered as moderately arousing. In this case, $Z_{ma} = 3.56$, $p = .0002$, $h = .44$. Finally, included as low arousal studies are those in which a weapon is present but no crime occurs or the central character carries a dangerous object (meat cleaver), not a prototypical weapon. For these seven studies, $Z_{ma} = .16$, $p = .44$, $h = .05$.

Confidence intervals computed for the above three groups show no overlap between the low arousal group [CI(95%) = $-.13$ to .23] and the moderate [CI(95%) = .27 to .61] or high arousal [CI(95%) = .30 to .46] groups. The narrowest interval is around the high arousal group.

Weapon Type

Assuming that the prototypical weapon is a gun, studies that use a gun ($N = 10$) were examined. This subset produces $Z_{ma} = 2.69$, $p = .0036$, $h = .14$. As a comparison, the only other weapon used in more than one test of the hypothesis, the meat cleaver, generated $Z_{ma} = .68$, $p = .25$, $h = .08$.

Lineup Type

The majority of tests used offender-present lineups ($N = 13$). For this group, $Z_{ma} = 2.82$, $p = .002$, $h = .12$. Five of the tests include a manipulated variable of offender present/absent lineups. In two of these tests, the interaction between the lineup type and weapon focus cannot be assessed (i.e., the variables are confounded); in the remaining three tests, one reports a significant interaction between the variables, such that the weapon focus effect occurred in the offender-present lineup only. The only test with a clear offender-absent lineup also produces a significant effect, $Z = 1.96$, $p = .03$, $h = .43$.

From a data set including all types of lineups it is evident that lower percent-

ages of correct identifications in the weapon-absent (control) condition are associated with larger effect sizes, $r(17) = -.44$, at a level approaching significance. Thus, it appears that scenarios (and more specifically, lineups) that produce low identification accuracy for subjects in general (i.e., control subjects) accentuate the weapon-focus effect.

Retention Interval

The amount of time between viewing of the event and subsequent lineup testing varied across studies from same-day testing to a 28-day delay. Comparison of same-day versus 2+ days of interval shows the following outcomes: 14 same-day tests produce $Z_{ma} = 2.34$, $p = .009$, $h = .12$; delayed lineups produce $Z_{ma} = 2.44$, $p = .007$, $h = .22$ ($N = 4$).

Published versus Unpublished Data

Four tests of the hypothesis were unpublished (Bothwell, 1991 [2]; Kramer, 1990; and Johnson, 1976). Kramer reports nonsignificant results; calculation of a Z score and an effect size for his data shows $Z_{ma} = .16$, $h = .04$. Johnson's data generate $Z_{ma} = 1.21$, $p = .11$, and $h = .35$; Bothwell's work produces $Z_{ma} = -.79$, $h = -.22$.

Subjects

Subjects were male and/or female undergraduate college students in all but one of the studies. O'Rourke et al. specifically addressed a concern of sample limitation by including a wide range of noncollege subjects from age 18 to 74, with results supporting the generalizability of eyewitness findings across age groups and subject populations.

DISCUSSION

Status of the Hypothesis

The data support the hypothesized weapon focus effect, with the combination of all tests producing a Z_{ma} of 3.40, $p = .0003$. This significant difference between weapon-present and weapon-absent conditions is apparent in the overall analysis and in many subset comparisons. The data also show that both dependent measures—lineup identification accuracy and feature accuracy—are sensitive to the weapon focus effect. The fail-safe N's calculated lend support to the credibility and generality of this sample of studies. The presence of a weapon does make a significant difference in eyewitness performance.

Although the overall effect size generated for lineup accuracy is not of great magnitude, it is well within a theoretically consistent range given that weapon absence or presence is only one of many variables that investigators recognize as influential in lineup identification accuracy. The greater effect size found for

WEAPON FOCUS

feature accuracy measures underscores the relevance of weapon focus in understanding eyewitness processes beyond the dependent measure of lineup identification.

The weapon focus effect has been found to be relatively robust across variations in stimulus presentation, experimental scenario, and experimenter and subject variables. The effect has been documented primarily with college student subjects, but also with noncollege populations. The effect occurs across experimental modes of "real life" enactments, video, and slide presentations; for experimental scenarios of high to moderate arousal; for both offender-absent and offender-present lineups; and with varying intervals of retention time. Most importantly, the effect is clearly discernable in those experimental paradigms that speak to the real-world issue at hand: situations in which a witness observes a threatening object play a central role in an event of short duration. Also it appears that the weapon focus effect is accentuated by longer intervals between the observed event and the lineup task and in low-optimality situations where the identification task is quite difficult.

The presence of a weapon during a crime is an "estimator" variable (Wells, 1978) in eyewitness performance. That is, the influence of this variable on an eyewitness's performance can only be estimated post hoc. Yet the data here do offer a rather strong statement: To not consider a weapon's effect on eyewitness performance is to ignore relevant information. The weapon effect does reliably occur, particularly in crimes of short duration in which a threatening weapon is visible. Identification accuracy and feature accuracy of eyewitnesses are likely to be affected, although, as previous research has noted (e.g., Pigott, Brigham, & Bothwell, 1990), there is not necessarily a concordance between the two. This information represents an important update in our state of knowledge regarding eyewitness accuracy, and as argued by Kassin et al. (1989), the weapon focus effect provides a good example of a phenomenon that may be viewed differently at this time compared to its status as a hypothesis just a few years before. There is now a convergence of data to support confidence in this effect.

Theoretical Implications

Past theoretical conjecture about the cause of the weapon focus effect has centered on two factors: arousal level and focus of attention. The present data show that the effect, as expected, is more pronounced in research scenarios that appear as real life to the subject. The effect is also present even in more artificial experimental settings when subjects view commission of a crime in which the weapon is clearly a threatening object. To the extent that moderate to high subject arousal levels may be inferred from these situations, it may be assumed that arousal is a correlate of a strong weapon focus effect.

Kramer et al.'s (1990) research, however, suggests that even in low-arousal situations a weapon can have an impact on feature accuracy. His data, which found the weapon focus effect to be dependent on the percentage of time the weapon was visible, do not rule out arousal as a causal factor, but suggest an explanatory emphasis on attentional processes. In an attempt to demonstrate

more directly the importance of focus of attention in this phenomenon, Loftus et al. (1987) employed a corneal reflection device. This tool allowed researchers to track eye movements of subjects, noting both number and duration of eye fixations. Their results provide direct evidence that eye fixation on a weapon is a critical correlate of reduced identification accuracy.

The importance of attention is also indirectly supported by the Cutler work. This research involved a weapon-absent condition in which the perpetrator's gun was in his pocket. Although it is difficult to isolate the reason for the lesser effect sizes produced by Cutler's scenario, perhaps the attention of these control group subjects was drawn to the pocket by what they inferred was there. Thus the attentional focus of control subjects may have been unintentionally similar to that of the experimental group.

Although focus of attention may be the critical mediating variable for the effect, it is also possible that the interaction of the two variables—narrowed focus of attention and high arousal—provides a situation that maximizes the potential for a weapon focus effect. As Loftus et al. suggest (1987; following Easterbrook, 1959), in a real-life crime situation, the narrowing of perceptual focus may be in fact accelerated by high arousal. A desirable goal for future research is to clarify the interactive effect of arousal and attention in the weapon focus phenomenon.

A variable related to both arousal and attention, crime scene complexity, could not be assessed with these data. It may be argued that real-life crime events include so many stimuli that the hypothesized weapon focus effect becomes irrelevant or insignificant in magnitude. The problem of ecological validity of these laboratory-based data can only be addressed here by noting that those researchers who have attempted to maximize complexity by adding additional bystanders or noise (e.g., Kramer, and Johnson) have not effectively eliminated the effect.

In sum, the weapon focus effect remains a worthwhile avenue for research. There is a need to more precisely identify the mechanics of the process in forensically relevant settings. At this point, however, the data provide evidence of its significant impact.

APPENDIX A: STUDIES INCLUDED IN THE META-ANALYSIS ($N = 12$)

Bothwell, R. K. (1991). *Eyewitness description accuracy*. Paper presented at the meeting of Southeastern Psychological Association, New Orleans, LA.

Cutler, B. L., & Penrod, S. D. (1988). Improving the reliability of eyewitness identification: Lineup construction and presentation. *Journal of Applied Psychology, 73*, 281–290.

Cutler, B. L., Penrod, S. D., & Martens, T. K. (1987a). The reliability of eyewitness identification: The role of system and estimator variables. *Law and Human Behavior, 11*, 223–258.

Cutler, B. L., Penrod, S. D., & Martens, T. K. (1987b). Improving the reliability of eyewitness identification: Putting context into context. *Journal of Applied Psychology, 72*, 629–637.

Cutler, B. L., Penrod, S. D., O'Rourke, T. E., & Martens, T. K. (1986). Unconfounding the effects of contextual cues on eyewitness identification accuracy. *Social Behavior, 1*, 113–134.

Johnson, C., & Scott, B. (1976). *Eyewitness testimony and suspect identification as a function of arousal, sex of witness, and scheduling of interrogation*. Paper presented at the American Psychological Association Annual Meeting, Washington, DC.

WEAPON FOCUS

Kramer, T. H. (1990). Personal communication, September, 1990.

Kramer, T. H., Buckhout, R., & Eugenio, P. (1990). Weapon focus, arousal and eyewitness memory. *Law and Human Behavior, 14,* 167–184.

Loftus, E. F., Loftus, G. R., & Messo, J. (1987). Some facts about "weapon focus." *Law and Human Behavior, 11,* 55–62.

Maass, A., & Kohnken, G. (1989). Eyewitness identification: Simulating the "weapon effect." *Law and Human Behavior, 13,* 397–408.

O'Rourke, T. E., Penrod, S. D., Cutler, B. L., & Stuve, T. E. (1989). The external validity of eyewitness identification research: Generalizing across subject populations. *Law and Human Behavior, 13,* 385–395.

Tooley, V., Brigham, J. C., Maass, A., & Bothwell, R. K. (1987). Facial recognition: Weapon effect and attentional focus. *Journal of Applied Social Psychology, 17,* 845–859.

APPENDIX B: SUMMARY OF STUDY CHARACTERISTICS

Study	Date	Effect size	Z	Total N	Arousal	Mode	Weapon	Interval
Bothwell	1991	−.43	−1.11	27	Low	Video	Gun	Same day
		.00	.00	28	Low	Video	Knife	Same day
Cutler and Penrod	1988	−.14	−.91	175	Moderate	Video	Gun	Same day
Cutler et al.	1987a	.29	1.32	165	Moderate	Video	Gun	Delay
Cutler et al.	1987b	.04	.34	290	Moderate	Video	Gun	Delay
Cutler et al.	1986	.00	.00	320	Moderate	Video	Gun	Same day
		.23	1.93	287	Moderate	Video	Gun	Delay
Johnson and Scott	1976	.35	1.21	48	High	Staged	Opener	—
Kramer	1990	.04	.16	62	Low	Slides	Cleaver	Same day
Kramer et al.	1990	.61	1.64	64	Moderate	Slides	Bottle	Same day
		.06	.25	64	Low	Slides	Cleaver	Same day
		.00	.00	32	Low	Slides	Cleaver	Same day
		.43	1.50	48	Low	Slides	Cleaver	Same day
		−.12	−.38	42	Low	Slides	Cleaver	Same day
Loftus et al.	1987	.67	1.93	36	Moderate	Slides	Gun	Same day
		.47	2.06	80	Moderate	Slides	Gun	Same day
Maass and Kohnken	1989	.43	1.96	86	High	Staged	Syringe	Same day
O'Rourke et al.	1989	.22	1.28	132	Moderate	Video	Gun	Delay
Tooley et al.	1987	.08	1.65	96	Moderate	Slides	Gun	Same day

REFERENCES

Beaman, A. L., Cole, C. M., Klentz, B., Preston, M., & Steblay, N. M. (1983). Fifteen years of foot-in-the-door research: A meta-analysis. *Personality and Social Psychology Bulletin, 9,* 181–196.

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences* (rev. ed.). New York: Academic Press.

Easterbrook, J. A. (1959). The effect of emotion on the utilization and organization of behavior. *Psychological Review, 66,* 183–201.

Frye v. United States, 293 F. 2d 1013 (D.C. Cir. 1923).

Hedges, L. V. (1984). Advances in statistical methods for meta-analysis. In W. H. Yeaton & P. M.

Wortman (Eds.), *Issues in data synthesis: New directions for program evaluation* (Vol. 24, pp. 25–42). San Francisco: Jossey-Bass.

Hunter, J. E., & Schmidt, F. L. (1990). *Methods of meta-analysis: Correcting error and bias in research findings.* Newbury Park, CA: Sage.

Kassin, S. M., Ellsworth, P. C., & Smith, V. L. (1989). The general acceptance of psychological research on eyewitness testimony: A survey of the experts. *American Psychologist, 44,* 1089–1098.

Kuehn, L. L. (1974). Looking down a gun barrel: Person perception and violent crime. *Perceptual and Motor Skills, 39,* 1159–1164.

Loftus, E. F. (1974). Reconstructing memory: The incredible eyewitness. *Psychology Today, 8,* 116–119.

Pigott, M. A., Brigham, J. C., & Bothwell, R. K. (1990). A field study on the relationship between quality of eyewitnesses' descriptions and identification accuracy. *Journal of Police Science and Administration, 17,* 84–88.

Rosenthal, R. (1978). Combining results of independent studies. *Psychological Bulletin, 85,* 185–193.

Rosenthal, R. (1984). *Meta-analytic procedures for social research.* Beverly Hills, CA: Sage.

United States v. Amaral, 488 F. 2d 1148 (9th Cir. 1973).

Wells, G. L. (1978). Applied eyewitness-testimony research: System variables and estimator variables. *Journal of Personality and Social Psychology, 36,* 1546–1557.

Wells, G. L. (1985). The eyewitness. In S. M. Kassin, & L. S. Wrightsman (Eds.), *The psychology of evidence and trial procedure* (pp. 43–66). Sage: Beverly Hills.

Yarmey, A. D., & Jones, H. P. T. (1983). Is eyewitness identification a matter of common sense? In S. M. A. Lloyd-Bostock & B. R. Clifford (Eds.), *Evaluating witness evidence: Recent psychological research and new perspectives* (pp. 13–40). New York: Wiley.

Eyewitness Accuracy Rates in Police Showup and Lineup Presentations: A Meta-Analytic Comparison[1]

Nancy Steblay,[2,5] **Jennifer Dysart,**[3] **Solomon Fulero,**[4] **and R.C.L. Lindsay**[3]

Meta-analysis is used to compare identification accuracy rates in showups and lineups. Eight papers were located, providing 12 tests of the hypothesis and including 3013 participants. Results indicate that showups generate lower choosing rates than lineups. In target present conditions, showups and lineups yield approximately equal hit rates, and in target absent conditions, showups produce a significantly higher level of correct rejections. False identification rates are approximately equal in showups and lineups when lineup foil choices are excluded from analysis. Dangerous false identifications are more numerous for showups when an innocent suspect resembles the perpetrator. Function of lineup foils, assessment strategies for false identifications, and the potential impact of biases in lineup practice are suggested as additional considerations in evaluation of showup versus lineup efficacy.

KEY WORDS: eyewitness; lineup; showup; meta-analysis.

An eyewitness to a crime quickly becomes a potentially critical factor in the apprehension and conviction of the perpetrator. Subsequent to a criminal event in which the perpetrator and witness are strangers, an identification procedure provides a memory test of the witness that can aid police in ascertaining whether a suspect is in fact the perpetrator. The most common police identification test procedures (Lindsay, 1999) are multiperson photo or live displays (lineups) and presentation of a single person to the witness (showup). In recent years, eyewitness researchers have identified flaws in police identification practices and have explored corrective avenues through comparative tests of alternative procedures. This line of research has contributed to the determination of best practices for obtaining and preserving eyewitness evidence (Technical Working Group for Eyewitness Evidence, 1999). The

importance of this applied research—and the techniques that have been developed for reducing the likelihood of false identification—is highlighted by dramatic news reports of DNA exoneration of convicted persons and the contribution of eyewitness identification errors in such cases (Wells et al., 2000).

Research to date has focused primarily on photo and live lineups, with less attention given to showup procedures despite evidence of their frequent use. Flowe, Ebbesen, Burke, and Chivabunditt (2001) report that showups were used for 55% of identifications conducted in 488 sampled cases between 1991 and 1995 in a large U.S. metropolitan area. McQuiston and Malpass (2001) document a showup use rate of 30% for identification attempts by police in El Paso County, Texas. Gonzalez, Ellsworth, and Pembroke (1993) enlisted the help of an Illinois detective to record all identifications (lineups and showups) in which he was involved over a designated period of time. Results from this field study indicated that 77% of identification tasks were showups. Thus, showups are a common and sometimes favored police identification procedure.

The small amount of research attention given to showups compared to lineups may be due to an expectation that a showup is simply an abbreviated lineup. A showup in fact may present a cognitive task quite similar to that of a lineup—and procedural recommendations for lineups (e.g., unbiased instructions) should logically be extended to showups. Alternately, a showup task may tap a slightly different cognitive strategy or set of situational influences, thus demanding a separate assessment of strengths and weaknesses. The correctness of either of these two positions is unclear at present, giving rise to the need for an empirical and evaluative comparison of showups with other identification techniques.

Researchers have empirically explored lineup formats and developed theoretical models of how lineup presentation is likely to affect witness decision-making. For example, lineups may be conducted either in simultaneous or sequential manner, and choice of format has been demonstrated to make a significant difference in level and type of ensuing errors (Steblay, Dysart, Fulero, & Lindsay, 2001; Wells et al., 1998). Simultaneous lineup presentation involves presentation of a group of photos or persons all at once to a witness, requiring the witness to decide if one of the displayed lineup members is the perpetrator. This technique allows a witness to compare lineup members and then to select the person who most closely resembles his or her memory for the culprit (i.e., a "relative judgment" strategy; see Wells, 1984). A simultaneous procedure will yield acceptable results when the perpetrator is in fact in the lineup, as the witness's comparison of the lineup members will often lead to the choice of the perpetrator as the closest match to memory. However, when the culprit is absent from the lineup, many witnesses continue to use the relative judgment strategy, resulting in an increase in the selection of an innocent lineup member, or "false alarm" (Steblay et al., 2001; Wells et al., 1994; Wells et al., 1998).

Lindsay and Wells (1985) proposed an alternative identification procedure, designed to restrict a witness's ability to use the relative judgment strategy. This technique, known as the sequential lineup, involves presentation of lineup members one at a time, requiring a yes/no identification decision for each member before the next one is shown. With this method, a witness must compare each lineup member to his/her memory of the culprit (i.e., the witness must make absolute judgments).

Accuracy Rates in Showup and Lineup Presentations

Lindsay and Wells found support for the superiority of the sequential strategy over the traditional simultaneous technique. Their results showed that the sequential and simultaneous techniques produced nearly equivalent correct identification rates in target-present lineups, yet the sequential technique produced close to a 25% lower false identification rate than the simultaneous method.

A recent meta-analysis comparing the effectiveness of the simultaneous and sequential lineup techniques confirmed what has been dubbed the "sequential superiority effect" (Steblay et al., 2001). The results support the reasoning of Lindsay and Wells (1985) that the sequential lineup forces eyewitnesses to use a more absolute judgment criterion rather than a relative judgment strategy. More specifically, participants in sequential lineup conditions were less likely to choose from the lineup, thus lowering their decision effectiveness for target present lineups but also reducing false identifications in target absent conditions. The reverse was true of the simultaneous lineup witness: An increased tendency to choose generated greater hit rates in the target present condition but also increased false identification errors by 23% in the target absent array, including a 200% increase in false identification of a designated innocent suspect. These outcomes reveal the complexity of lineup presentation issues. For example, the 15% increase in accuracy found for target-present simultaneous lineups appears desirable, but may in fact be due to calculated guesses. Also, under conditions approximating real-life, benefits of the target-present simultaneous presentation were found to diminish whereas the target-absent advantages of sequential lineups remained stable.

Given that the showup identification procedure is a one-photograph technique requiring only one "yes–no" judgment, it should logically provide the benefit of absolute judgment (i.e., fewer false alarms). This line of thinking suggests that an eyewitness faced with a showup will be less likely to choose than when viewing a lineup, thereby reducing both correct and false identifications. One might further predict that a showup would be at least as effective as a sequential lineup and superior to a simultaneous format.

However, one of the benefits of a lineup—either simultaneous or sequential—is that there is some protection for the innocent suspect in the presence of lineup foils. An unreliable eyewitness or absence of the true perpetrator in the lineup can be signaled by a witness's selection of a foil. The showup does not offer such protection. In addition, while the eyewitness to a lineup can correctly assume that there will be more than one choice in the task, a showup is understood by the eyewitness to be a single opportunity to identify the perpetrator. The showup reveals police suspicions about the single suspect, and the witness is aware that only one person will be shown. Thus the procedure may be considered an "inherently suggestive one" (Lindsay & Wells, 1980; Phillips, McAuliff, Kovera, & Cutler, 1999). Indeed, using the same term, the United States Supreme Court (*Stovall v. Denno*, 1967; *United States v. Wade*, 1967) and many state courts (*Bradley v. State*, 1980; *Commonwealth v. Carter*, 1979; *Holden v. State*, 1979) have acknowledged that showups are suggestive. This suggestiveness may affect outcomes by generating more choosing from showups than lineups. Offering some support to this speculation, Behrman and Davey (2001) found that in actual criminal cases, 76% of witnesses in showup circumstances made identifications, whereas only 48% of witnesses in photo lineups did so. If choosing

is increased, the showup procedure may generate an increase in both correct and incorrect choices or simply make witnesses more likely to identify an innocent suspect as the perpetrator without affecting the rate of correct choices. In either case, the benefit gained by an absolute judgment strategy may be balanced or negated by pressure to choose and the fact that identification errors cannot be spread across foils (known errors). Following this logic, one might predict that a showup would be particularly dangerous for innocent suspects and thus less desirable as an identification procedure.

This project is an extension of the past work that compared sequential and simultaneous lineup formats (Steblay et al., 2001). Meta-analysis will be used to compare showup to lineup presentation strategies. Most showup researchers have tested the hypothesis that a one-person showup increases the likelihood of misidentification compared to a full lineup. A recent survey of experts (Kassin, Tubb, Hosch, & Memon, 2001) found that 74% of respondents considered that finding to be reliable, and 85% reported that their opinion was based on published, peer reviewed scientific research.

A preliminary review of past research highlights four intriguing points relevant to this survey finding. First, there is very little available research that explicitly compares showup to lineup performance. Only eight articles, with 12 tests, have been located after extensive investigation. Second, the available research on showup identifications has yielded inconsistent results. A quick tally shows four reports of the negative impact of showups (Lindsay, Pozzulo, Craig, Lee, & Corber, 1997; Wagenaar & Veefkind, 1992; Yarmey, Yarmey, & Yarmey, 1994, 1996), one that suggests that showups produce more accurate identifications (Beal, Schmitt, & Dekle, 1995), and two reports (Dekle, Beal, Elliott, & Huneycutt, 1996; Gonzalez, Ellsworth, & Pembroke, 1993;) that indicate equivocal or no difference in decision outcomes. This variability in study outcome highlights a third issue: Interpretation of outcome is somewhat a function of the dependent measure of interest—positive identifications, choosing rates, or false identifications. Outcomes of prior lineup research suggest that exploration of multiple dependent measures will provide a more complete picture of this complex phenomenon. A final point, as noted above, is that reasonable extrapolation from existing theory and empirical work may lead one to opposing predictions about eyewitness choosing and accuracy levels in showups compared to lineups. For these reasons, a summary report of showup performance is necessary.

A central purpose of meta-analysis is to search the data for any underlying pattern, a consistent display of an effect despite surrounding noise. Subsequent exploration of theoretical and methodological variables that moderate an effect often highlights and clarifies nuances of a complex phenomenon. Essential commonality of hypothesis is critical to the studies that make up a meta-analysis, yet diversity in method addressing that hypothesis typically affords access to more complete knowledge. Despite the small number of empirical studies available on the topic of showup performance, this meta-analysis is anticipated to provide useful supplementary knowledge to our growing understanding of eyewitness performance. This expectation is based on the high quality of studies available—seven of the eight are published—and the attention within these studies to relevant theoretical questions

and issues central to eyewitness identification practice. The studies present a desirable variety of approaches and samples. For example, Wegenaar and Veefkind (1992) provide two studies, the first a laboratory test utilizing a slide sequence stimulus and including 548 citizen subjects; their second experiment tapped a more realistic scenario involving a staged crime and a full week delay between crime and identification task during which the college subjects did not know that they would be called back to a lineup task. Gonzales, Ellsworth, and Pembroke (1993) in their first study staged a classroom incident that involved their participants in a cross-racial identification task. In a second study, these researchers explored variations of crime (theft in a restaurant) and modality (video) as well as a change to same-race identification in the context of lineup foils of high, medium, or low similarity to the perpetrator. Lindsay et al. (1997) explored subject sample differences (preschool, school-age, and college students) in a noncrime event with a lineup selection procedure that used every member of the target-absent lineup in the target-absent showup condition. This diversity of researchers' approaches provides the potential for meaningful exploration of the parameters of showup/lineup performance even within the small sample.

Consistent with the Kassin et al., survey of experts, this meta-analysis begins with the primary hypothesis that a showup will lead to increased false identifications compared to a lineup. Additional complexities of the showup-lineup comparison also will be explored, with expectation of less dramatic differences between showups and lineups in target-present scenarios. The research will compare showups and lineups on three primary outcomes: Rates of overall correct identification decisions; correct identifications of perpetrators from target present arrays; and misidentification errors from target absent arrays. The evaluation of misidentification errors is more complicated than at first may appear. In the case of a perpetrator-absent array, a clear comparison can be made in the laboratory between the rates of correct rejection from showups versus lineups. It is more difficult to compare false identification rates. False identification and false positive selection rates are identical for showups (only one choice is available), while in a perpetrator-absent lineup, the innocent suspect may be chosen (false identification) or a foil selection is possible. Care will be taken to distinguish between these two choices (Lindsay et al., 1997).

Additionally, the predicted tendency for false identification to occur more often with showups may depend on how the innocent suspect is selected. Many, but not all, researchers select innocent suspects based on their similarity to the confederate. The result when innocent replacements are determined by other means is not clear and there may be insufficient data at this time to test this effect. However, it is hoped that the available data allows testing of three outcomes of target-absent arrays: Correct rejection rates for showups versus lineups; false identification rates of designated innocent suspects selected on the basis of similarity to the confederate for showups versus lineups; and false identification rates of designated innocent suspects that do not resemble the confederate for showups versus lineups. The specific goals of this meta-analysis are (1) to generate a quantitative and theoretical summary of research findings that compare showup and lineup performance, (2) to ascertain the state of the research literature, and (3) to provide direction for future research efforts.

METHOD

Sample

A computer search of the PsycINFO database provided an initial sample of studies relevant to the hypothesis. Direct contact with lineup researchers provided access to additional tests and more complete data. In order to be included in the sample, the experimental study must have compared showup to lineup performance and provided a statistical test of the relationship between presentation format and identification accuracy. Both sequential and simultaneous lineup formats were included as lineup tests. Multiple dependent measures of accuracy were available in the sample, and the review incorporated performance frequencies of the following: (1) overall correct decisions, collapsed across target-present and target-absent presentation (correct identifications plus correct rejections); For target-present presentations, (2) correct identifications, (3) false rejections, and (4) choice of a foil (a known error); For target-absent formats, (5) correct rejections, (6) identification of any foil, and (7) identification of a designated innocent suspect or target.

Eight papers were located (seven published and one unpublished), providing 12 tests of the hypothesis. The data set included studies completed between 1977 and 2002, representing 3013 participants. Both male and female participants were included in all tests. Sample sizes ranged from 59 to 565, with a mean of 251.08. The set includes data from 1127 community residents (41%), 1320 undergraduates (44%), and 459 (15%) children.

Study Characteristics

Methodological and theoretical variables were coded as part of the data set. Methodological variables included researcher, year of publication, source (published or unpublished), number of hypothesis tests per study, sample size, subject sex, sample makeup (children, undergraduate students, adult sample, mixed), lineup size, lineup mode (live, photo, video), design (between-subject, within-subject), type of crime (robbery, vandalism, non-criminal), event stimulus (video, live, slides), and procedural blinds (double-blind, no double-blind). Variables of more theoretical import included time of delay between event and identification task (immediate, 2 days to 1 week), number of perpetrators, race and gender of perpetrator, inclusion of a verbal description task (present, not present), instruction (biased, unbiased) lineup construction (biased, unbiased), lineup type (sequential, simultaneous), lineup construction strategy (match-to-description, match-to-target), choice of target replacement (best match to target, rotation of foils), and exposure time in seconds.

All 12 studies provided a between-subject design, a lineup of size 6, unbiased lineup instructions, and a single perpetrator. Time of exposure to the perpetrator ranged from 2 to 90 s, with a mean of 57 s. Only one test specifically reported use of a double-blind procedure (Wegenaar et al., Experiment 2, 1992); the remaining articles included no comment regarding double-blind precautions.

Accuracy Rates in Showup and Lineup Presentations

Two authors (JD & NS) independently recorded data from each paper, and then compared information to check for oversights. Design variables were coded by a team of student researchers. These codes were derived directly from the papers, with minimal interpretation necessary. Multiple coders were employed simply to assure that available information was recorded correctly. Thus ultimate agreement among coders was 100%.

Statistics

Following the work of Rosenthal (1991), the Pearson correlation coefficient r was used as the measure of effect size. The mean effect size for a group of hypothesis tests is referred to in subsequent discussion simply as r. A meta-analytic Z (Z_{ma}) was calculated by combining Z-scores of individual tests of the hypothesis using the Stouffer method (Rosenthal, 1991). This method produces an overall probability level associated with the observed pattern of results. A fail-safe N (N_{fs}) was calculated to estimate the number of additional tests averaging null results that would be needed in order to bring the significance level attained through the meta-analysis to a value larger than .05.

RESULTS

Twelve tests of the hypothesis were available to examine the status of the effect, that is, that lineup presentation fosters better eyewitness performance than does a showup format. Positive r and Z values denote support of this hypothesis. Negative r and Z values indicate results in the opposite direction, that is, that subjects in the showup condition performed with greater accuracy than subjects in the lineup condition. Comparisons are considered as one-tailed tests.

Overall Frequency of Correct Decisions

The first pass through the data set was to ascertain the overall level of correct identification decisions by eyewitness subjects. These figures, from 12 tests, represent the frequency of correct identifications in target-present presentations plus correct rejections in target-absent presentations. Showup presentation produced a mean of 69% correct decisions; lineups generated 51% correct decisions, a significant difference, $Z_{ma} = -9.31$, $p < 0001$, $N_{fs} = 372$, with an effect size $r = -.18$, favoring the showup. This calculation, however, does not distinguish type of error committed. That analysis requires consideration of a critical moderator variable in lineup research: whether the perpetrator is present in or absent from the array (e.g., see Steblay, 1997 and Steblay et al., 2001). The next calculations attempt to assess the impact of showup versus lineup presentation for target-present and target-absent presentations separately (see Table 1).

Table 1. Identification Performance: Showup Versus Lineup

	N	Showup (%)	Lineup (%)	r	Seq (%)[a]	Sim (%)[a]
Overall correct decisions	12	69	51[b]	−.18	56	48
Target present display						
Correct ID	12	47	45	−.02	35	50
Miss	12	53	55		65	50
False rejection	10	58	34[b]	.26	46	26
Foil ID	10	—	24		19	24
Target absent display						
Correct rejection	11	85	57[b]	−.32	72	49
Miss	11	15	43		28	51
Miss minus Foil IDs	5	—	16			
False identifications of innocent suspect (minus foil IDs)	3	23	17	.07	09	27

[a]Sequential and Simultaneous Lineups, from Steblay et al., 2001.
[b]$Z_{ma} > 1.65$, $p < .05$.

Frequency of Choosing[6]

A primary empirical question is whether presentation format affects choosing behavior of witnesses. Collapsed across target-present and target-absent conditions, the data indicate that the witness is twice as likely to choose from a lineup as from a showup (54% vs. 27%). In target-present conditions, 71% of subjects viewing a lineup made a choice from the array (either a correct or a foil ID) and 46% of showup subjects made a choice, in this case, a correct ID. (These data are from a subset of nine tests with the necessary information.) For 11 studies with target absent displays, lineups again produced a higher choosing rate: 43% versus 15%, lineups versus showups, respectively. Accuracy of the choosers is addressed in subsequent sections.

Decision-Making in Target-Present Conditions

For the eyewitness presented with a target-present showup or lineup, two outcomes are possible: Correct identification of the perpetrator (a "hit") or a failure to identify (a "miss"). A miss can take the form of an incorrect rejection of the display, an "I don't know" (DK) response, or in the case of a lineup, selection of a foil. The data (Table 1) demonstrate that correct identification is slightly more likely in the target-present showup presentation than in the lineup format, $Z_{ma} = -1.38$, $p = .08$, $r = -.02$ (based on $N = 12$), with a 2% performance advantage (47% vs. 45%, showup versus lineup, respectively). Inversely, the overall *miss* or error rate of showups compared to lineups is 53% versus 55%. Effect sizes for correct identifications compared between lineups and showups are displayed on Table 2.

---

[6]"Choosers" in this analysis represent those participants who select a member of the array, correctly or incorrectly. Nonchoosers are those who reject the lineup, correctly or incorrectly, or report that they "don't know." This definition differs from the Gonzales et al. (1993) "choosers" who were defined as those who were confident enough to "decide" as to the presence or absence of the perpetrator in the array. The current analysis extracted and used the Gonzales data consistent with our definition of choosers.

Accuracy Rates in Showup and Lineup Presentations

Table 2. Stem and Leaf Display of Effect Sizes r. Target Present Presentation: Correct Identifications

Stem	Leaf
.7	
.6	
.5	
.4	
.3	
.2	3,4
.1	6,8
.0	4,7
−.0	1,7,9
−.1	
−.2	2
−.3	5,7
−.4	
−.5	
−.6	
−.7	

A more precise breakdown of error type can be determined in a subset of 10 tests. False rejections when the target is present (including DKs) are significantly fewer in the lineup condition, $Z_{ma} = 7.41$, $p < .0001$, $N_{fs} = 132$, $r = .26$ (58% vs. 34%, showups versus lineups). Foil identifications account for the remaining misses in the lineup condition: Twenty-four percent of subjects in the target-present lineup condition chose a foil, a known error.

Another way to view this outcome is to consider only "choosers." Showup presentations generate a significantly lower rate of choosing than do lineups, 46% versus 71% in target-present conditions. In a target-present showup, just making a choice assures a hit (100% true positive identification), while lineup choices allow for distribution across foils, thus potentially reducing true positives. Target-present lineup accuracy for choosers is 64%, a significantly lower hit rate compared to showups, $Z_{ma} = −10.18$, $p < .0001$, $N_{fs} = 336$, $r = −.42$, $N = 9$. Thus, lineups produce higher choosing (71%) with a lower hit rate (64%), and showups produce lower levels of choosing (46%) with a higher hit rate (100%). Overall in target-present presentations, the showup and lineup will produce approximately the same results (46% vs. 45% correct identifications).

Decision Making in Target-Absent Conditions

Two outcomes are possible for an eyewitness confronted with an identification task that does not include the perpetrator: correct rejection of the array (which may be in the form of "I don't know") or false identification. In this case, showups produced a significantly higher level of correct rejections compared to lineups (85% vs. 57%), $Z_{ma} = 11.76$, $p < .0001$, $N_{fs} = 552$, $N = 11$, $r = −.32$. Inversely, the showup produced 15% errors, compared to 43% in the lineup (see Table 1). Effect sizes for the comparison of correct rejections between lineups and showups are displayed in Table 3.

Table 3. Stem and Leaf Display of Effect Sizes r. Target Absent Presentation: Correct Rejections

Stem	Leaf
.7	
.6	
.5	
.4	
.3	
.2	
.1	
.0	
−.0	
−.1	7,9
−.2	4,6,6,9
−.3	8
−.4	1,2,5,8
−.5	
−.6	
−.7	

The 15% showup error rate represents a "dangerous" error: Identification of an innocent suspect as the perpetrator. In the case of a lineup, this same error may occur. In addition, however, a lineup may generate a foil selection—a known error. The 43% lineup error rate mentioned above includes both false identifications and foil identifications. Teasing apart these two types of error produces a more precise indicator of dangerous lineup error rate. Five tests allowed separation of foil and suspect choices (Dekle et al., 1996; Gonzales et al., 1993, Experiments 1 and 2; Yarmey et al., 1994, 1996). These five tests appear representative of the larger data set, in that showup vs. lineup error rates of the five average 15% and 41%, respectively (compared to the full sample rates of 15% and 43%).

There are two ways to consider lineup error rates in these five tests. The 41% error rate can be divided into foil (31%) and suspect (10%) identifications. Thus, 10% of lineup decisions result in "dangerous" false IDs, compared to 15% of showup decisions. An alternative procedure is to subtract foil choices from the analyses (reducing the overall number of subjects in the analysis). With foil choices excluded, lineups generate an 84% correct rejection rate and 16% "dangerous" false identification rate, virtually the same as showups (85% correct rejection, 15% false ID), $r = -.03$.

In this subset of five studies, the error rate of choosers from the target-absent showup is 100% (a choice is automatically a false identification), but a smaller percentage of subjects are choosers, 16%. The target-absent lineup-choosing rate is 44%, with a 25.2% false identification rate (based on five tests). The overall error rate is 11% versus 16%, showups versus lineups.

Three of the research teams (Dekle et al., 1996; Yarmey et al., 1994, 1996) have further explored the perpetrator-absent scenario by planting a suspect in the lineup or showup who closely matches the description of the perpetrator. This person becomes an "innocent suspect" in the perpetrator-absent lineup and showup. As mentioned above, overall error rates are higher in lineups (23% vs. 45%, showups vs. lineups, respectively). However, dangerous false identification in these cases is

higher in showups than lineups, 23% versus 10%. With lineup foil choices excluded from analysis, this "dangerous" false identification rate, showups to lineups, is 23% versus 17%, $Z_{ma} = 1.57$, $p = .06$, $r = .07$.

Comparison of Showups to Sequential and Simultaneous Formats

A related question for this investigation of showup/lineup performance is the comparison of outcomes for showups versus simultaneous and sequential lineup presentations. Steblay et al., (2001) reported that participants in the sequential lineup condition are less likely to make a lineup choice. This lower choosing rate results in false rejection errors if the target is in the lineup and reduces false identification errors if the target is absent. The reverse pattern occurs with the simultaneous lineup: An increased tendency to choose favors this participant if the target is indeed present while increasing false identification errors, particularly of a designated innocent suspect, in a perpetrator absent lineup. The two right-most columns of Table 1 report figures for sequential and simultaneous lineups from the Steblay et al. meta-analysis.

Showup/lineup comparisons in this study echo some patterns seen in the sequential/simultaneous comparison. Like sequential lineups, showups produce fewer choices, and in doing so lead to more false rejections in the target-present condition and more correct rejections in the target-absent condition. An area of difference, however, is apparent for correct identifications. In the current study, showups and lineups are approximately equal in true positive identifications, while in the earlier work, simultaneous lineups produced significantly better eyewitness performance (15%) than sequential lineups. The majority of lineups represented in the current data set are of simultaneous format. Of the 12 tests included in this meta-analysis, three used the sequential lineup format for some or all of the lineup data, and two of the three were studies repeated from the Steblay et al., (2001) meta-analysis. Eliminating these three from the data set produces a lineup comparison group that consists of only simultaneous format. Effect sizes are similar to the overall group, .03 and -.32 in target-present and absent conditions, respectively.

An additional difference emerges for the showup/lineup comparison when the planted innocent suspect in a target-absent lineup is considered. Steblay et al. report that simultaneous lineups generated three times more false identifications than sequential lineups. In this data set showups produce more false identifications, 12% versus 5%, when compared to simultaneous lineups, however this outcome should be considered tentative, as it is based on only two tests.

Moderator Variables

As the stem and leaf displays indicate, there is essential commonality in target absent performance, rs ranging from $-.17$ to $-.48$. Table 4 illustrates the consistency of effect sizes in target absent conditions and suggests minimal impact of moderator factors.

The Target present column of Table 4 for the most part replicates the earlier analyses—indicating just small differences between showup and lineup performance regardless of moderator variable. However, target present performance indicates

Table 4. Effect Size Analysis by Moderator Variables

	r (N)	
Variable	Target present: Correct identification	Target absent Correct rejection
Sample		
Preschool to kindergarten	$-.36\ (3)^a$	$-.27\ (2)^a$
Children 8–10 years	$.00\ (1)$	$-.35\ (1)^a$
Children 11–15 years	$.03\ (1)$	$-.31\ (1)^a$
Undergraduates	$.07\ (8)$	$-.34\ (7)^a$
Citizens 18–65 years	$-.16\ (2)^a$	$-.22\ (2)^a$
Lineup construction		
Unbiased	$-.02\ (11)$	$-.31\ (10)^a$
Biased toward foil	$.07\ (1)$	$-.41\ (1)^a$
Lineup/showup type		
Photo	$-.04\ (11)$	$-.31\ (10)^a$
Live	$.23\ (1)$	$-.48\ (1)^{a,b}$
Delay between event and identification task		
Immediate (no delay)	$-.03\ (8)$	$-.33\ (7)^a$
2 days to 1 week	$.10\ (3)$	$-.34\ (3)^a$
Verbal description		
Yes	$.00\ (9)$	$-.31\ (9)^a$
No	$-.05\ (3)$	$-.40\ (2)^a$
Event stimulus mode		
Live	$-.03\ (6)^a$	$-.30\ (5)^a$
Slides or transparencies	$-.04\ (5)$	$-.36\ (5)^a$
Video	$.18\ (1)$	$-.26\ (1)^a$
Event		
Robbery/theft	$.03\ (7)$	$-.36\ (7)^a$
Smashed equipment	$.24\ (1)$	$-.17\ (1)$
Noncriminal event	$-.16\ (4)^a$	$-.28\ (3)^a$
Perpetrator gender		
Male	$.08\ (6)$	$-.34\ (5)^a$
Female	$-.12\ (6)^a$	$-.31\ (6)^a$
Publication status		
Published	$-.03\ (10)^a$	$-.30\ (9)^a$
Not published	$.03\ (2)$	$-.43\ (2)^a$

[a] $Z_{ma} > 1.65$, $p < .05$.
[b] The one test (Gonzales et al., 1993) represented in this category is also the only test involving a cross-racial identification. Thus the individual impact of these two factors cannot be separated.

lack of consistency on two levels. First, Table 4 reveals some variability across subsets within a variable, e.g., age of sample produces effect sizes ranging from $-.36$ to $+.07$. Second, not evident on Table 4 is the actual variability in effect sizes that underlies average effect sizes hovering around zero. It is important to note that these average outcomes conceal tests with both negative and positive signs (sometimes lineup, sometimes showup superiority). Target present outcomes merit attention, to explore conditions under which effects are most pronounced or constrained. Unfortunately, analysis of moderator variables in this data set is limited by the small number of tests available and the uniformity of some design components across studies.

In target present conditions, showup performance is elevated for child participants, non-criminal stimulus events, and when the perpetrator was female. As these factors are confounded within studies, it is not possible to separate out their impact. There is also an increased showup superiority associated with adult (citizen)

populations. Somewhat superior levels of correct identification in lineups are associated with undergraduates, longer time delays between event and identification task, male perpetrators, and one study in which subjects saw a staged crime in their classroom ("smashed equipment"). Lineup performance is also better in the one study that used a live lineup and a cross-racial identification, factors that could not be separated for analysis.

DISCUSSION

When overall identification decisions are tabulated, showups produce an accuracy advantage over lineups (69% vs. 51%). This initial result is qualified by subsequent analyses. As anticipated, a consideration of specific subject choices provides a more complete picture. Correct identification (hit) rate within the context of a target-present condition is nearly identical for the two types of procedures: Approximately 46% of witnesses shown either a lineup or a showup correctly identified the perpetrator when he or she was present. False suspect identification rates in a target-absent display are also approximately equal between showups and lineups, at about 16%. Analysis of error type provides a reason for the discrepancy between the initial overall showup accuracy advantage and the hit and false identification outcomes just described. Witnesses who choose from a target-absent lineup produce more errors, but also divert their erroneous choices across foils. If foil identifications are categorized as errors, the error rate of lineups increases dramatically. Alternately, when foil identifications instead are folded into the category of nonidentification of the suspect, showup and lineup outcomes converge. Overall, the results present surprising commonality in outcome between presentation formats, and—specific to target-absent displays—an apparent contradiction of the ambient knowledge that showups are more dangerous for innocent suspects than are lineups. Additional factors will also inform a comparison of showups versus lineups. These involve our understanding of the function of lineup foils, assessment of false identifications, and the potential for biases in lineup practice.

Lineup Foils

Following the logic of the above discussion, the role of lineup foils is a first consideration. Used effectively, a lineup will serve two purposes: To determine whether a suspect is in fact the perpetrator observed by the witness and to assess the reliability of the witness. A foil selection suggests unreliable witness memory and discredits the witness rather than the suspect. The lineup witness who selects a foil may rightly be considered an unreliable source for subsequent identification evidence. Only witnesses who reject the lineup by choosing no one may be considered a credible source in a subsequent identification task. On the other hand, the showup witness has no foil options. A witness who rejects the showup retains police trust as a reliable witness, even in the case of a false rejection of a target-present showup. This is potentially dangerous in the face of a subsequent identification attempt with a new and innocent suspect. Therefore, if foil choices are considered useful indications that witnesses are

willing to identify innocent people, lineups have an advantage. (Current data show a 24% foil identification rate in target-present lineups and 31% foil identification rate in target-absent lineups.)

The use of foils to detect an unreliable witness is of particular interest when one considers very young children. A substantial amount of research literature has examined the eyewitness reporting accuracy of children compared to adults. Wells, Wright, and Bradfield (1999) summarize this literature specific to lineup performance, by pointing out that "The primary condition for concern in eyewitness identification from lineups and photospreads is the condition in which the actual perpetrator is not present..." (p. 60). Dekle et al., (1996) similarly report the literature as showing that while child witnesses make correct identifications from target-present lineups at the approximate level of adults, children are more likely than adults to choose someone from a target-absent lineup (a false ID), even when warned directly that the perpetrator may not be in the lineup. In this data set, children exhibit better target-absent performance for showups than lineups, as did adults. In target-present arrays young children (preschool and kindergarten) also performed significantly better on showups than lineups, a finding that deserves attention in future work. Part of that analysis must be a differentiation of target identification (choices of the target among multiple responses from an individual child) versus correct identification (a single and correct identification). Young witnesses have a tendency to make multiple choices from a lineup, thus impeaching their own testimony. As noted by Lindsay et al. (1997), the reliability of eyewitness identification is thus "seriously compromised by the tendency for children to guess" (p. 401). This impeachment through multiple choices cannot occur in a showup, thus perhaps affecting the appearance of better showup performance.

The counterargument to this framework for consideration of identification tasks—that foil choices that discredit the witness represent a problem for police and the solution of crimes—and the choice of whether such witnesses should be considered a source of identification evidence is a policy, not empirical, issue.

False Identification Rates

Experts who responded to the Kassin et al. (2001) survey expressed specific concern with false suspect identification rates of showups. This is reasonable, given that only false identifications lead to the risk of false accusation and wrongful conviction. As reported above, this meta-analysis has identified approximately equal false identification rates from showups versus lineups (16%), deriving that figure by direct tabulation of errors in target-absent lineups. Previous research teams have attempted to estimate false identification rates through two other means. First, the overall rate of false positive choices can be divided by the nominal size or number of people examined during the identification procedure to generate an expected false identification rate. This approach (Lindsay, Pozzulo, Craig, Lee, & Corber, 1997) is based on the assumption that the innocent suspect is no more likely than any other lineup member to resemble the criminal if the lineup has been constructed based on matching foils to the description of the criminal provided by the witness. The lineups in this meta-analysis had a nominal size of 6, thus the expected false

Accuracy Rates in Showup and Lineup Presentations

identification rate would be 43%/6 = 7.2%. This compares to the substantially higher rate of 15% for showups. Based on the sequential lineup meta-analysis, 6-person sequential lineups have an expected false identification rate of 5.33% (Steblay et al., 2001). If this approach is accepted, then showups do represent a greater risk of false identification.

A related measure, the diagnosticity ratio, is employed by some researchers to establish utility of a lineup procedure (Wells & Lindsay, 1980). The advantage to this perspective is that police know whether or not the witness chooses someone but do not know if the identification procedure is criminal-present versus criminal-absent. For showups 42% of witnesses choose the suspect in the present condition and 15% in the absent condition, generating a diagnosticity ratio of 2.80 (Wells & Lindsay, 1980). For lineups, 42% of witnesses choose the suspect from the present lineup and 16% from the absent lineup, generating a diagnosticity ratio of 2.6.

A second approach to estimating false identification rates is to designate a specific, criminal-absent lineup member as the innocent suspect. The rate at which the designated individual is identified is considered the false identification rate. Within this tradition, two approaches have been taken. One assigns the innocent suspect role randomly (or perhaps haphazardly) to the six lineup members, while the other and more common approach assigns the role of innocent suspect to the absent lineup member deemed most similar to the criminal. In the current data set, the innocent suspect not explicitly selected based on similarity to the criminal, produced almost identical rates of false identification (15 and 16%, showups to lineups, respectively). If the innocent suspect was selected based on similarity to the criminal, showups generated more choices of that designated innocent suspect (23% vs. 17% respectively). The showup then may be equivalent in risk for an innocent suspect only to the extent that the innocent suspect does not strongly resemble the true criminal.[7]

Bias

The vulnerability of an innocent suspect who matches the description of the perpetrator illustrates a third factor for consideration. It is reasonable to wonder if other means of influencing a decision criterion, for example, clothing, instruction, and foil biases (Lindsay, Wallbridge, & Drennan, 1987; Lindsay & Wells, 1980; Malpass & Devine, 1981; Steblay, 1997) may increase false identifications differentially for showups and lineups. Although foil bias will not be a factor with showups,

---

[7]Differences in practice are apparent in the studies at two points during formation of a target-absent array. At a first point, researchers identify appropriate lineup foils. Six of the hypothesis tests in this data set indicate a "match to general description" strategy for foil determination, and five tests use a "match to target" method (one test not reported). A match-to-target was typically used as a means to construct high/medium/low foil similarity for exploration of that variable. Analysis of witness error rates indicate small differences associated with construction strategy: Match-to-description produced 13% showup and, 44% lineup errors; Match-to-target produced 17% showup and 43% lineup errors.
As a second step, researchers decide on a target replacement for the target-absent showup or lineup. In four tests, the authors used a strategy that essentially rotated the lineup foils through the position of target replacement for the target-absent showup. For six tests, the target replacement was the foil most resembling the perpetrator. Again, witness error rates differed slightly based on strategy: Use of a target match produced 15% showup errors, 42% lineup errors; Rotation of foils produced 12% showup errors, 44% lineup errors.

clothing and instruction bias may be influential. All studies in this set used nonbiased instructions, thus the effect of instructional bias on showup performance remains an open question. The high rate of correct rejection for criminal-absent showups may reflect reactance to the suggestiveness of the procedure. Combining biased instructions with the showup procedure may therefore result in two distinct outcomes: biased instructions may increase reactance and thus further decrease false positive choices, or biased instructions could alleviate witness concerns that the procedure is biased and dramatically increase false identifications. Current research being conducted on this issue favors the latter explanation (Dupuis, Dysart, & Lindsay, 2001).

Clothing bias may be of particular concern with showups as the procedure is used shortly after the crime and frequently in the field rather than at police stations. Apprehension of suspects for showups is generally based on the combination of a match to the description provided by the witness and proximity to the crime. As a result, suspects will generally be wearing clothing that resembles the witness' description of clothing worn by the criminal during the crime. The fact that the showup generally occurs shortly after the crime may further convince witnesses that the suspect is unlikely to be innocent. They may ask themselves "How many people can there be in this area that look like that and are wearing clothes like that?" The less time between the crime and the showup, the stronger this intuition may be. Dysart, Dupuis, and Lindsay (2001) have recently found strong evidence of clothing bias with showups, indicating that the type of clothing worn by the perpetrator may interact with other factors, such similarity of the innocent suspect. Although the results from this study are compelling, the data did not include a lineup comparison. We are left with an incomplete picture of showup vulnerability to bias, but reason to speculate that several known lineup biases may influence showups as well.

Theoretical and Future Research Considerations

Significantly lower levels of choosing behaviors for witnesses presented with a showup versus a lineup suggest that, even though decision outcomes may be similar, differential decision processes may be attendant to the two identification formats. Given this, it is appropriate to ascertain what we can about witness reliability and strategy from these available data.

As discussed by previous researchers (Lindsay & Wells, 1985), an absolute decision process is desirable, particularly as a means to reduce false identifications. The lower level of choices in showup conditions may be construed as an indication that subjects are in fact using, at least more so than in the lineup condition, an absolute judgment strategy. The increased rejection rates (false rejections in target present conditions and correct rejections in target-absent conditions) suggest that showup subjects have attained some benefit of absolute judgment, perhaps due to a showup's similarity to a "one-person" sequential lineup. On the other hand, it is apparent that lineup foil options provide a deflection of error away from an innocent suspect and a valuable vehicle to identify the unreliable witness. These benefits help to equalize lineup and showup performance, at least under the rather favorable conditions of these studies.

Evaluation of this data set must include concern regarding the small number of studies available. In fact, one key outcome of the investigation is to alert the research community to the paucity of data and to the need for more deliberate attention to showups. The analyses exposed performance variability yet to be explored in target-present scenarios. The small number of hypothesis tests in this data set deterred analysis of some potentially fruitful variables, as exemplified by the Gonzales et al., study. These authors included in their method two relevant and intriguing components—cross-racial identification and use of a live lineup—that could not be independently examined because they are confounded within that study and not available in other tests. Therein lies direction for future research.

Finally, the showup's potential for suggestibility—which worries legal professionals and eyewitness experts—is evidenced in this data set, although in a small number of studies. The data currently available leave us with residual concern regarding potential dangers of showups and with a strong appreciation of the need for research that will specifically address showup accuracy under realistic conditions comparing competent practice with biased procedure.

REFERENCES

References marked with an asterisk indicate studies included in the meta-analysis.

*Beal, C. R., Schmitt, K. L., & Dekle, D. J. (1995). Eyewitness identification of children: Effects of absolute judgments, nonverbal response options, and event encoding. *Law and Human Behavior, 19,* 197–216.

Behrman, B., & Davey, S. (2001). Eyewitness identification in actual criminal cases: An archival analysis. *Law and Human Behavior , 25,* 475–491.

Bradley v. State, 264 SD.E.2d 332 (Ga. App. 1980).

Commonwealth v. Carter, 414 A.2d 369 (Pa. Super. Ct. 1979).

*Dekle, D. J. (1993). *The effects of testing delay and composite viewing on identification accuracy in lineups and showups*. Unpublished doctoral dissertation, Dartmouth College.

*Dekle, D. J., Beal, C. R., Elliott, R., & Hunneycutt, D. (1996). Children as witnesses: A comparison of lineup versus showup identification methods. *Applied Cognitive Psychology, 10,* 1–12.

Flowe, H., Ebbesen, E., Burke, C., and Chivabunditt, P. (2001, July). *At the scene of the crime: An examination of the external validity of published studies on line-up identification accuracy*. Paper presented at the American Psychological Society Conference, Toronto, Canada.

*Gonzalez, R., Ellsworth, P. C., & Pembroke, M. (1993). Response biases in lineups and showups. *Journal of Personality and Social Psychology, 64,* 525–537.

Holden v. State, 602 P.2d 452 (Alaska 1979).

Kassin, S. M., Tubb, V. A., Hosch, H. M., & Memon, A. (2001). On the general acceptance of eyewitness research: A new survey of the experts. *American Psychologist, 56,* 405–416.

Lindsay, R. C. L. (1999). Applying applied research: Selling the sequential lineup. *Applied Cognitive Psychology, 13,* 219–225.

*Lindsay, R. C. L., Pozzulo, J. D., Craig, W., Lee, K., & Corber, S. (1997). Simultaneous lineups, sequential lineups, and showups: Eyewitness identification decisions of adults and children. *Law and Human Behavior, 21,* 391–404.

Lindsay, R. C. L., & Wells, G. L. (1980). What price justice? Exploring the relationship of lineup fairness to identification accuracy. *Law and Human Behavior, 4,* 303–313.

Lindsay, R. C. L., & Wells, G. L. (1985). Improving eyewitness identifications from lineups: Simultaneous versus sequential lineup presentation. *Journal of Applied Psychology, 70*(3), 556–564.

McQuiston, D., & Malpass, R. (2001, June). *Eyewitness identifications in criminal cases: An archival study*. Paper presented at the fourth biennial meeting of the Society for Applied Research in Memory and Cognition, Kingston, Ontario, Canada.

Phillips, M. R., McAuliff, B. D., Kovera, M. B., & Cutler, B. L. (1999). Double-blind photoarray adminsitration as a safeguard against investigator bias. *Journal of Applied Psychology, 84*(6), 940–951.

Rosenthal, R. (1991). *Meta-analytic procedures for social research*. Newbury Park, CA: Sage.

Steblay, N. (1997). Social influence in eyewitness recall: A meta-analytic review of lineup instruction effects. *Law and Human Behavior, 21,* 283–297.

Steblay, N., Dysart, J., Fulero, S., & Lindsay, R. C. L. (2001). Eyewitness accuracy rates in sequential and simultaneous lineup presentations: A meta-analytic comparison. *Law and Human Behavior, 25,* 459–473.

Stovall v. Denno, 388 U.S. 293 (1967).

Technical Working Group for Eyewitness Evidence. (1999). *Eyewitness evidence: A guide for law enforcement [booklet]* . Washington, DC: United States Department of Justice, Office of Justice Programs, Document Number NCJ-178240.

United States v. Wade, 388 U.S. 218 (1967).

*Wagenaar, W. A., & Veefkind, N. (1992). Comparison of one-person and many-person lineups: A warning against unsafe practices. In F. Losel, D. Bender, & T. Bliesner (Eds.), *Psychology and law: International perspectives* (pp. 275–285) Berlin: Walter De Gruyter.

Wells, G. (1984). The psychology of lineup identifications. *Journal of Applied Social Psychology, 14,* 89–103.

Wells, G. L., & Lindsay, R. C. L. (1980). On estimating the diagnosticity of eyewitness nonidentifications. *Psychological Bulletin, 88,* 776–784.

Wells, G. L., Malpass, R. S., Lindsay, R. C. L., Fisher, R. P., Turtle, J. W., & Fulero, S. M. (2000). From the lab to the police station: A successful application of eyewitness research. *American Psychologist, 55,* 581–598.

Wells, G., Small, M., Penrod, S., Malpass, R., Fulero, S., & Brimacombe, E. (1998). Good practice recommendations for lineups and photospreads. *Law and Human Behavior, 22,* 603–647.

Wells, G. L., Wright, E. F., & Bradfield, A. L. (1999). Witnesses to crime: Social and cognitive factors governing the validity of people's reports. In R. Roesch, S. D. Hart, & J. R. P. Ogloff (Eds.), *Psychology and law: The State of the discipline* (pp. 53–87). Kluwer Academic/Plenum: New York.

*Yarmey, A. D., Yarmey, A. L., & Yarmey, M. J. (1994). Face and voice identifications in showups and lineups. *Applied Cognitive Psychology, 8,* 453–464.

*Yarmey, A. D., Yarmey, A. L., & Yarmey, M. J. (1996). Accuracy of eyewitness identification in showups and lineups. *Law and Human Behavior, 20,* 459–477.

Research Article

Postidentification Feedback Affects Real Eyewitnesses

Daniel B. Wright and Elin M. Skagerberg

University of Sussex, Brighton, United Kingdom

ABSTRACT—*Many studies of simulated eyewitness situations have shown that under certain laboratory conditions, people's confidence about their identifications predicts their accuracy, but that their reported confidence can be affected by telling them that they chose the suspect. In this study, eyewitnesses (n = 134) to real crimes took part in lineups at an identification suite in the United Kingdom and were asked questions about their memory both before and after they were told whether they had identified the suspect or a filler. Before the eyewitnesses were told whether they had identified the suspect or a filler, their responses to several questions reliably differentiated between those who identified the suspect and those who identified a filler. In addition, responses to the memory questions were affected by telling the eyewitnesses whether or not they had identified the suspect. These results show that postidentification feedback affects real eyewitnesses and highlight the importance of recording meta-memory variables before an eyewitness discovers whether he or she has identified the suspect.*

Differentiating accurate from inaccurate eyewitness identifications is a goal of many police investigators and juries, and has become a vibrant area of psychological research. Under certain conditions, how confident a person is can be a moderate to good predictor of his or her accuracy, but in other conditions, confidence is a poor predictor of accuracy (Brewer & Wells, 2006). Sporer, Penrod, Read, and Cutler's (1995) meta-analysis of 30 studies showed that, on average, when subjects made an identification, their confidence was a good predictor of their accuracy ($r = .37$, 95% confidence interval: .32–.42). However, there was more variability among the individual correlations than would be expected if all these studies had measured a single correla-

tion value. The 95% credibility interval (i.e., that interval expected to include approximately 95% of all the population correlations for different situations) was from .20 to .55. Given this interval, one might conclude that the correlation between confidence and accuracy for actual eyewitness identifications is likely to be somewhere between .20 and .55. Unfortunately, this would not be a valid conclusion precisely because the results of the meta-analysis show that there is not one single association for all situations, which means it is vital that the studies sampled be representative of all potential situations of interest (Wright, 2006). Although many of the authors of the studies included in the meta-analysis attempted to mimic real crime situations, the set of studies is not representative of the entire population of interest.

One of the most important findings of the past decade of research on eyewitness testimony is that a person's confidence in his or her testimony can be affected by what other people say. Wells, Douglass (née Bradfield), and other researchers (e.g., Bradfield, Wells, & Olson, 2002; Dixon & Memon, 2005; Wells & Bradfield, 1998) have conducted several laboratory studies demonstrating that responses to meta-memory questions like "how difficult was it for you to make an identification?" are affected by telling respondents that they identified the suspect or that they identified an innocent filler. In a meta-analysis of the size of this postidentification-feedback effect in different situations, Douglass and Steblay (2006) found the effect is stronger for measures of certainty than for other measures, such as responses regarding the eyewitness's view of the culprit and his or her memory in general.

Douglass and Steblay (2006) also found a larger effect for responses providing confirming evidence than for responses providing disconfirming evidence. This difference may have been due to the fact that most of the studies they sampled used target-absent lineups, which made the identification task difficult and often impossible.[1] If people make a choice from a six-

Address correspondence to Daniel B. Wright, Psychology Department, University of Sussex, Brighton, BN1 9QH, UK, e-mail: danw@sussex.ac.uk.

[1] The film Wells and his colleagues often use for their studies on postidentification feedback can be viewed on the Web at http://www.psychology.iastate.edu/faculty/gwells/theeyewitnesstest.html. The film we often use for such studies can be viewed at http://www.sussex.ac.uk/Users/danw/Elinweb.htm.

person lineup and are told that they have chosen the suspect, they may be surprised because the task was difficult; this surprise may result in a larger postidentification-feedback effect than is found in people who are told that they have chosen a filler. The situation may be different with real lineups, however. The police officers we have spoken with have said that eyewitnesses seem more surprised when told that they have identified a filler than when told that they have identified the suspect. Therefore, Douglass and Steblay's finding may not generalize to actual lineups.

In the present study, our goal was to take these findings, two of the most important in eyewitness research, out of the laboratory and see if they occur with real eyewitnesses taking part in real lineups ("identification parades" in British English). There have been several archival studies of lineups (e.g., Valentine, Pickering, & Darling, 2003; Wright & McDaid, 1996), but the eyewitnesses in these studies were not asked for any meta-memory judgments. The most relevant archival study was conducted by Behrman and Richards (2005; see also Behrman & Davey, 2001). They took the verbal descriptions often given by eyewitnesses (phrases like "absolutely certain," as opposed to "not quite sure") and used these to calculate measures of confidence. They found eyewitnesses who identified fillers had lower confidence than those who identified the suspect. There has been no similar study looking at postidentification feedback.

A difficulty analyzing data from real lineups is not knowing if the suspect is in fact the culprit. If much is known about the case, it is possible to estimate the likelihood that the suspect is guilty independent of the identification, as Behrman and Davey (2001) did. It is also possible to look at courtroom decisions, but these are affected by eyewitnesses' identifications, so that they do not provide an independent measure of accuracy. Because of the anonymity requirements for our data collection, it was not possible to match the responses from our eyewitnesses with more detailed case information. Therefore, we had to make some assumptions about the performance of eyewitnesses. The first assumption was that identification of a filler was an error. Estimating how often eyewitnesses are correct when they identify the suspect and when they identify no one in the lineup is more difficult. Penrod (2003) used Behrman and Davey's data to estimate how often eyewitnesses make correct identifications. He estimated that about 75% of the time when no identification is made, the culprit is in the lineup. He also estimated that about 85% of the time when the suspect is identified, he or she is guilty, although about 15% of the time, this correct identification is actually a guess. These estimates are based on many assumptions, including an assumption about the overall proportion of lineups in which the suspect is guilty, which varies by jurisdiction (Wells & Olson, 2002). If a jurisdiction conducts lineups only when the suspect is almost certainly guilty, then very seldom will an innocent suspect be identified. However, if lineups are used mainly to exonerate suspects who the police think are probably innocent, a much larger proportion of

innocent suspects will be identified. Penrod's estimates provide only a rough guide, but it seems reasonable to assume that most identifications of suspects are correct.

Our first aim in this study was to discriminate suspect from filler identifications in real police lineups and examine if the size of the differences between groups on the meta-memory variables was similar to that found in laboratory studies comparing accurate and inaccurate identifications (Sporer et al., 1995). Our second aim concerned the malleability of responses to the meta-memory questions and the postidentification-feedback effect. We asked eyewitnesses six questions, although here we focus on the question, "How difficult was it for you to figure out which person in the identification parade was the person who committed the crime?" because this question is the one most associated with confidence ($r = .74$; Wells, Olson, & Charman, 2003, Table 3) and was shown to have a large effect in Douglass and Steblay's (2006) meta-analysis.

Our study took place in the United Kingdom, where lineup procedures differ in several important ways from those in the United States (Kebbell, 2000). Almost all lineups in the United Kingdom are now conducted in specialist lineup suites (which are separated from the rest of the police station and operated by their own personnel), using video files of the suspect and eight or more fillers (Valentine et al., 2003).[2] Valentine and Heaton (1999) described the video procedures in more detail and showed that video lineups are more fair than traditional live lineups. At the identification suite that participated in our study, the police officer constructing the lineup inputs some basic descriptive information about the suspect into a computer, and the computer searches a database of approximately 20,000 faces to produce a large set of potential fillers matching this description. The officer then chooses a much smaller subset for possible use in the lineup. The suspect's lawyer is given the opportunity to be involved in constructing the final lineup. The entire lineup is stored on a CD and shown to eyewitnesses when they become available. Thus, when there are multiple eyewitnesses, the same lineup is used for each.

The identification suites in the United Kingdom follow a code of practice that was outlined in the 1984 Police and Criminal Evidence Act (Home Office, 2005). One rule is that eyewitnesses have to look at each person twice before making any decision. Although Weber, Brewer, Wells, Semmler, and Keast (2004) found that the time to make an identification is negatively correlated with accuracy, because of this rule, the relation between response time and accuracy is more difficult to evaluate in the United Kingdom than in the United States.

---

[2]Suspects are not required to be video-recorded, but in almost all cases they agree to the recording. If they do not, the police are allowed to use photographs taken when the suspect was first detained, and a photo spread is used instead of a video lineup. A photo spread is also used when the suspect's appearance has changed since the time of the crime.

METHOD

Ethical and Practical Issues

In the typical eyewitness laboratory study, anonymity and confidentiality are maintained out of professional courtesy. It is unlikely anybody would be interested that a particular subject gave a particular answer. In the present study, ethical concerns were more serious because our data could be subpoenaed if lawyers felt these data could help their case. Therefore, we took safeguards so that it would be difficult to ascertain an individual's identity either from the data file or from the response sheets. We recorded only each eyewitness's identification during the lineup and whether the eyewitness was a victim or a witness. The eyewitness identifier used in the data file did not correspond to when the lineup took place. Finally, the written response sheets required eyewitnesses only to circle six numbers, so it would be difficult for a handwriting expert to identify an eyewitness from his or her response sheet. Ironically, one of the main conclusions from this study is that it would be good, in the future, to make meta-memory ratings available to the courts.

In laboratory studies, the researcher is free to ask subjects almost any meta-memory question. In the present study, more care was needed both because the eyewitness was often in a fragile state (e.g., the victim of an attack would likely be emotional immediately after having to think about the crime) and also because some questions were inappropriate for legal reasons. Taking into account the sensitivity of many eyewitnesses, we did not ask any eyewitnesses to rape or murder cases to participate. We examined all other cases individually to determine if the eyewitness might be in a particularly sensitive state.

After much discussion, we decided not to ask the eyewitnesses directly about their confidence. The words *confidence* and *certainty* have become part of legal jargon, and we felt that asking a question using the word *confidence* could prompt a lawyer involved in one of the cases to want access to our data. This would jeopardize the ethics (participants' anonymity) of the study, so this word was avoided. A question about how difficult the identification was served as a proxy for measuring confidence. We did not ask eyewitnesses whether they would be willing to testify in court, as Wells and Bradfield (1998) had, because the eyewitnesses could feel that their responses were legally binding.

The study was conducted in collaboration with the Sussex police and with their approval. The study received ethical clearance from the Psychology Ethics Committee at the University of Sussex.

Sample

The lineups occurred at the Sussex Identification Suite in Brighton, United Kingdom, where almost all lineups in Sussex occur. During the study, there were no murder cases. There were rape cases, but we did not ask the eyewitnesses in these cases

to participate. There were no other cases in which we felt the eyewitness would have been negatively affected by taking part in our study. When an eyewitness took part in more than one identification parade, we used data from the first parade only. In total, 136 eyewitnesses were approached, and 134 agreed to take part. Two thirds were victims.

Procedure

Before each lineup, the eyewitness arrived at the identification suite and was brought to a prelineup waiting room with a witness support officer and a guardian, if applicable. A civilian employee of the suite arrived, explained the lineup procedure, and took the eyewitness to the parade room, where a police officer was waiting. The police officer, who was a full-time employee of the suite and not connected with the case, showed the eyewitness the computer display system and emphasized that the culprit might or might not be in the lineup. The police officer usually knew which person in the lineup was the suspect, a problem that we discuss at the end of this article. The eyewitness was then shown the lineup in accordance with the Police and Criminal Evidence Act (PACE). If the eyewitness made an identification, the police officer did not say whether the suspect was identified.

The eyewitness was then taken by the civilian employee to a postlineup waiting room, where the witness support officer, the researcher, and a guardian, if applicable, were waiting. The researcher asked if the eyewitness would take part in the study. If the eyewitness agreed, he or she was randomly allocated to either Group A or Group B. Participants in Group A were asked the following three questions and responded using scales from 1 through 10 (verbal anchors of the scale are listed after each question):

- *View*: "How good a view did you have of the culprit during the crime?" (scale from *poor view* to *very good view*)
- *Time*: "Sometimes people are able to make an identification very quickly. Other times people have to think about it for a while. How quickly were you able to make your identification?" (scale from *immediately* to *had to think about it for a very long time*)
- *Face*: "Some people have good memories for the faces of strangers who they have only seen once, some people have poor memories. In general, do you have a good memory for people you have only seen once?" (scale from *poor memory* to *very good memory*)

Participants in Group B were asked three different questions, which also had scales from 1 through 10:

- *Attention*: "How much attention did you pay to the culprit during the crime?" (scale from *no attention* to *my complete attention*)

- *Ease*: "How difficult was it for you to figure out which person in the identification parade was the person who committed the crime?" (scale from *very difficult* to *very easy*)
- *Event*: "Some people have good memories for everyday events like going to the supermarket; some people have poor memories of these events. In general, do you have a good memory for everyday events?" (scale from *poor memory* to *very good memory*)

The questions were presented to the eyewitnesses on a response sheet, but the researcher was also there to read the questions if the eyewitness did not feel comfortable reading them. One eyewitness did not speak English, so the translator who was there with him for the lineup translated the questions and his responses.

Next, the police officer came into the room and told the eyewitness the outcome of the identification (i.e., whether he or she had identified the suspect or a filler). The researcher then had the eyewitness answer the three questions that he or she had not answered previously. Afterward, the eyewitness was debriefed and thanked. Eyewitnesses were not paid for participating.

The six meta-memory questions used are based on those used by Wells and Bradfield (1998). We chose them so that all eyewitnesses were asked about their encoding circumstances (view and attention questions), about their identification (time and ease questions), and about their general memory ability (face and event questions) both before and after discovering the lineup outcome. Using a design in which eyewitnesses are asked questions both before and after discovering the lineup outcome

makes it possible to collect more data from each eyewitness, but tends to decrease the size of the postidentification effect (Wells & Bradfield, 1998, Experiment 2).

Some eyewitnesses made no identification. The questions were not designed for them, so their responses are not considered here.

RESULTS AND DISCUSSION

Of the 134 eyewitnesses, 21% made no identification, 21% identified a filler, and 58% identified the suspect. These percentages are different from those obtained in past archival studies in the United Kingdom (i.e., Slater, 1994; Valentine et al., 2003; Wright & McDaid, 1996), which yielded percentages of approximately 40, 20, and 40%, respectively, but are closer to Behrman and Davey's (2001) percentages of 26, 24, and 50%. These percentages vary by jurisdiction because of differences in how much evidence is required to hold a lineup.

Less than 1% of responses were missing—in most cases, because the participant marked two responses. These scores were excluded only from analyses involving the variable with the missing values. The scores of the time variable were reversed. Thus, higher scores corresponded to better view, quicker identification, more attention, more easy identification, and better memory, and we expected that all items would be positively correlated with each other, which they were.

The top half of Table 1 shows results for eyewitnesses' responses prior to being told whether they identified the suspect.

TABLE 1

Responses to the Six Meta-Memory Questions, Before and After the Eyewitness Was Told the Outcome

| | Person identified as the suspect | | | | Difference between responses of subjects who identified the filler and subjects who identified the suspect | | | |
| | Filler | | Suspect | | | | | |
Question	M	SD	M	SD	95% confidence interval	t (df)	p_{rep}	r_{pb}
					Before discovering outcome			
View	6.75	2.93	8.32	2.26	(0.03, 3.16)	1.97 (51)	.87	.27
Time	4.91	2.51	7.88	2.86	(1.07, 4.87)	3.14 (51)	.97	.40
Face	6.00	1.91	7.47	2.19	(0.07, 2.86)	2.11 (53)	.89	.28
Attention	6.56	2.45	8.15	2.32	(0.14, 3.03)	2.21 (48)	.91	.30
Ease	4.88	2.73	7.76	2.85	(1.17, 4.60)	3.39 (48)	.98	.44
Event	6.38	1.96	7.79	1.82	(0.28, 2.56)	2.51 (48)	.94	.34
					After discovering outcome			
View	5.75	2.59	8.38	2.02	(1.28, 3.98)	3.92 (48)	.99	.49
Time	4.69	2.15	6.91	3.22	(0.67, 3.78)	2.88 (42.08)	.96	.34
Face	4.88	1.45	7.68	2.04	(1.66, 3.95)	4.92 (48)	.99	.58
Attention	5.17	2.55	8.11	2.21	(1.47, 4.49)	3.97 (51)	.99	.49
Ease	2.75	1.66	8.60	2.08	(4.54, 7.17)	8.96 (53)	.99	.78
Event	6.75	2.93	8.07	1.96	(−0.60, 3.24)	1.47 (13.86)	.76	.25

Note. In cases in which Levene's test for equality of standard deviations was significant, the *t* values, degrees of freedom, p_{rep} values, and confidence intervals were adjusted. The p_{rep} statistic is calculated from the *p* value (Killeen, 2005). A *p* value of .05 produces a p_{rep} of .88.

The correlations (r_{pb}) were all within the credibility interval reported by Sporer et al. (1995) and in Cohen's (1988) terms represent medium-sized effects. All the effects were in the predicted direction, with eyewitnesses who identified suspects reporting having better views and paying more attention to the culprit, making their decision more quickly, and finding the task easier than eyewitnesses who identified fillers. Eyewitnesses who identified suspects also reported having better memory for faces and events. This difference in reported memory may have arisen because in fact they did have better memory (and therefore did not choose fillers) or because they temporarily believed this, having just made an identification with more confidence than the eyewitnesses who chose fillers. The effects were fairly reliable, with p_{rep}s of .87 or higher (meaning there is at least an 87% probability that if the study were repeated, the observed effects would be in the same direction; Killeen, 2005).

The bottom half of Table 1 shows the same statistics for questions asked after the eyewitnesses were told whether they chose the suspect. The statistics showed the same pattern, and overall the effects were larger, with correlations in Cohen's (1988) medium to large range. The increased effect size was expected from the postidentification-feedback effect.

We examined the postidentification-feedback effect by conducting a 2 (choice: suspect, filler) × 2 (time of questioning: before discovering outcome, after discovering outcome) between-subjects analysis of variance for each of the meta-memory questions. As the means in Table 1 show, scores generally went down for participants who had chosen a filler and up for participants who had chosen the suspect. Given that Douglass and Steblay (2006) found a particularly large effect for measures of certainty, we focus on the ease question. Responses to this question showed a main effect of choice, $F(1, 101) = 65.68$, $p_{\text{rep}} = .99$, $\eta_p^2 = .39$, with people who chose the suspect thinking the task was simpler than those who chose a filler, and a nonsignificant effect, overall, for whether the question was asked before or after participants discovered the identity of the suspect, $F(1, 101) = 1.42$, $p_{\text{rep}} = .69$, $\eta_p^2 = .01$. The most important finding of this study is the significant interaction between choice and time of questioning, $F(1, 101) = 7.55$, $p = .006$, $p_{\text{rep}} = .96$, $\eta_p^2 = .07$. People who chose a filler showed a downward shift, from 4.88 to 2.75, or approximately 1 standard deviation. People who chose the suspect showed an upward shift, from 7.76 to 8.60, or approximately a third of a standard deviation. It is necessary to be cautious in comparing these effects because many responses of eyewitnesses who identified the suspect were near 10, so there was a ceiling effect.

Comparing means shows that overall there was a downward shift for participants who chose a filler and an upward shift for those who chose a suspect. However, this comparison does not indicate whether the shift varied with confidence (e.g., whether among eyewitnesses who chose the suspect, the least confident showed the greatest shift). Penrod (2003) estimated that about 15% of correct suspect identifications are what he called guesses

and that another 15% of suspect identifications are incorrect. It is a particular concern when eyewitnesses identify the suspect but find the task difficult, because these identifications are the most likely to be errant. To examine whether this group was particularly affected by postidentification feedback, we estimated the mean response for each decile (each 10%) using the Harrell-Davis estimator (Harrell & Davis, 1982). Figure 1 shows the differences between the values before participants discovered whether they identified the suspect and the values after participants discovered whether they identified the suspect, what is called the shift function (Wilcox, 1997). The graph shows that among eyewitnesses who identified the suspect, there was relatively little postidentification-feedback effect for people near the top of the ease scale, but for the lowest 20 to 30% of the scale, the shift was about 2 or 3 units, or approximately 1 standard deviation. Thus, people who without feedback would have responded with a 4 or 5 on the scale would respond with scores of 6 to 8 after feedback. The shift for eyewitnesses who chose fillers also depended on prediscovery ease ratings. There was little effect for those who found the task difficult. However, for those on the high end of the ease scale, the disconfirming feedback had a very large effect, such that people who without feedback would have given a rating of 8 or 9 would choose 4 or 5 after feedback. The size of these effects must be viewed with some caution because of both ceiling and floor effects.

SUMMARY AND RECOMMENDATIONS

Responses to each of six meta-memory questions discriminated between eyewitnesses who chose the suspect and those who chose a filler. This was true both before and after the eyewitnesses were told the outcome of their identification and has important forensic implications. Finding out whether they had identified the suspect or a filler affected eyewitnesses' responses to the ease question. It is important to note, however, that the police officer conducting the lineup was usually aware of which person in the lineup was the suspect. Although the officer did not tell the eyewitnesses the outcome, it would have been virtually impossible for even the most careful officer not to provide some verbal and nonverbal cues (see chapters in Sebeok & Rosenthal, 1981). Consequently, the effect sizes for the association between the meta-memory variables and lineup choice before finding out the lineup result may be inflated and the estimates for the postidentification-feedback effect deflated. Thus, the postidentification-feedback effect may be even larger than we report. We agree with other researchers (e.g., Wells et al., 1998) that double-blind procedures should be used for lineups.

Our main recommendations concern assessing confidence and other meta-memory variables. On the basis of a wealth of laboratory studies, Wells and his colleagues (Wells et al., 1998, 2000; Wells & Olson, 2003) have recommended recording confidence after eyewitnesses make an identification, but before they discover the outcome of the identification. In this study, the
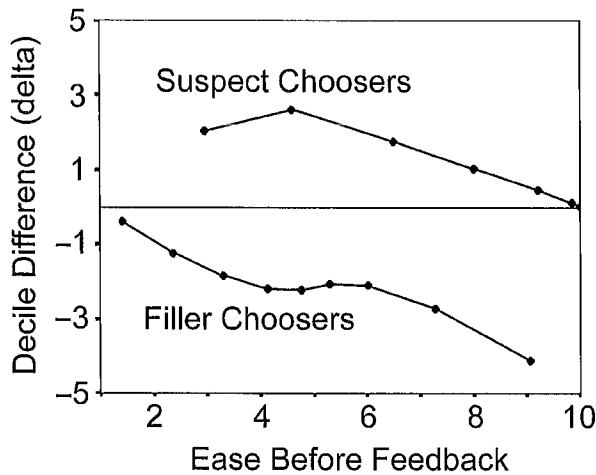
Fig. 1. Change (delta) in judged ease of the identification for each decile of responses to the initial ease question. Change was calculated by subtracting the estimated mean judgment before feedback from the estimated mean judgment after feedback. Results are shown separately for eyewitnesses who chose a suspect and those who chose a filler during the lineup.

meta-memory judgments of real eyewitnesses prior to finding out the outcome of the lineup were moderate predictors of accuracy (if one assumes most of the identifications of suspects were accurate). Thus, we agree that recording confidence and other variables at this point is of diagnostic value and could help jurors and other individuals to assess the reliability of an identification.

We also found that the postidentification-feedback effect occurs with real eyewitnesses. Telling eyewitnesses that they did or did not pick the suspect affects how difficult they think the task was. It is a particular concern that the effect was strong for those people who chose the suspect and found the task relatively difficult (Fig. 1). According to Penrod's (2003) analysis, approximately 30% of suspect identifications are guesses, and about half of these guesses are identifications of an innocent suspect. Confirming feedback may mask any apparent uncertainty that might help to differentiate accurate from inaccurate identifications. It is critical that any recording of eyewitnesses' confidence be done prior to when they discover the outcome of the lineup; otherwise, their confidence will be contaminated by the postidentification feedback.

REFERENCES

Behrman, B.W., & Davey, S.L. (2001). Eyewitness identification in actual criminal cases: An archival analysis. *Law and Human Behavior, 25*, 475–491.

Behrman, B.W., & Richards, R.E. (2005). Suspect/foil identification in actual crimes and in the laboratory: A reality monitoring analysis. *Law and Human Behavior, 29*, 279–301.

Bradfield, A.L., Wells, G.L., & Olson, E.A. (2002). The damaging effect of confirming feedback on the relation between eyewitness certainty and identification accuracy. *Journal of Applied Psychology, 87*, 112–120.

Brewer, N., & Wells, G.L. (2006). The confidence-accuracy relationship in eyewitness identification: Effects of lineup instructions, foil similarity and target-absent base rates. *Journal of Experimental Psychology: Applied, 12*, 11–30.

Cohen, J. (1988). A power primer. *Psychological Bulletin, 112*, 155–159.

Dixon, S., & Memon, A. (2005). The effect of post-identification feedback on the recall of crime and perpetrator details. *Applied Cognitive Psychology, 19*, 935–951.

Douglass, A.B., & Steblay, N.M. (2006). Memory distortion in eyewitnesses: A meta-analysis of the post-identification feedback effect. *Applied Cognitive Psychology, 20*, 859–869.

Harrell, F.E., & Davis, C.E. (1982). A new distribution-free quantile estimator. *Biometrika, 69*, 635–640.

Home Office. (2005). *Police and Criminal Evidence Act 1984 (s.60(1)(a), s.60A(1) and s.66(1)). Codes of Practice A-G. 2005 edition.* London: The Stationary Office.

Kebbell, M.R. (2000). The law concerning the conduct of lineups in England and Wales: How well does it satisfy the recommendations of the American Psychology-Law Society? *Law and Human Behavior, 24*, 309–315.

Killeen, P.R. (2005). An alternative to null-hypothesis significance tests. *Psychological Science, 16*, 345–353.

Penrod, S. (2003, Spring). Eyewitness identification evidence: How well are witnesses and police performing? *Criminal Justice Magazine, 54*, 36–47.

Sebeok, T.A., & Rosenthal, R. (Eds.). (1981). *The Clever Hans phenomenon: Communication with horses, whales, apes, and people.* New York: New York Academy of Sciences.

Slater, A. (1994). *Identification parades: A scientific evaluation.* London: Police Research Group, Home Office.

Sporer, S.L., Penrod, S., Read, D., & Cutler, B. (1995). Choosing, confidence, and accuracy: A meta-analysis of the confidence-accuracy relation in eyewitness identification. *Psychological Bulletin, 118*, 315–327.

Valentine, T., & Heaton, P. (1999). An evaluation of the fairness of police line-ups and video identifications. *Applied Cognitive Psychology, 13*, S59–S72.

Valentine, T., Pickering, A., & Darling, S. (2003). Characteristics of eyewitness identification that predict the outcome of real lineups. *Applied Cognitive Psychology, 17*, 969–993.

Weber, N., Brewer, N., Wells, G.L., Semmler, C., & Keast, A. (2004). Eyewitness identification accuracy and response latency: The unruly 10–12 second rule. *Journal of Experimental Psychology: Applied, 10*, 139–147.

Wells, G.L., & Bradfield, A.L. (1998). "Good, you identified the suspect": Feedback to eyewitnesses distorts their reports of the witnessing experience. *Journal of Applied Psychology, 83*, 360–376.

Wells, G.L., Malpass, R.S., Lindsay, R.C.L., Fisher, R.P., Turtle, J.W., & Fulero, S. (2000). From the lab to the police station: A successful application of eyewitness research. *American Psychologist, 55*, 581–598.

Wells, G.L., & Olson, E.A. (2002). Eyewitness identification: Information gain from incriminating and exonerating behaviors. *Journal of Experimental Psychology: Applied, 8*, 155–167.

Wells, G.L., & Olson, E.A. (2003). Eyewitness identification. *Annual Review of Psychology, 54*, 277–295.

Wells, G.L., Olson, E.A., & Charman, S.D. (2003). Distorted retrospective eyewitness reports as functions of feedback and delay. *Journal of Experimental Psychology: Applied, 9*, 42–52.

Wells, G.L., Small, M., Penrod, S., Malpass, R.S., Fulero, S.M., & Brimacombe, C.A.E. (1998). Eyewitness identification procedures: Recommendations for lineups and photospreads. *Law and Human Behavior, 22*, 603–647.

Wilcox, R.R. (1997). *Introduction to robust estimation and hypothesis testing*. San Diego, CA: Academic Press.

Wright, D.B. (2006). Causal and associative hypotheses in psychology: Examples from eyewitness testimony research. *Psychology, Public Policy, and Law, 12*, 190–213.

Wright, D.B., & McDaid, A.T. (1996). Comparing system and estimator variables using data from real line-ups. *Applied Cognitive Psychology, 10*, 75–84.